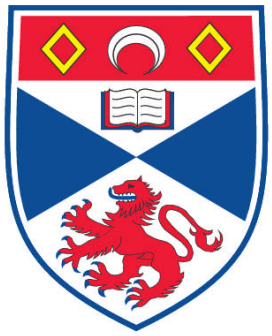
THE DESATURASE GENE FAMILY: AN EVOLUTIONARY STUDY OF PUTATIVE SPECIATION GENES IN 12 SPECIES OF *DROSOPHILA*

Maria C. Keays

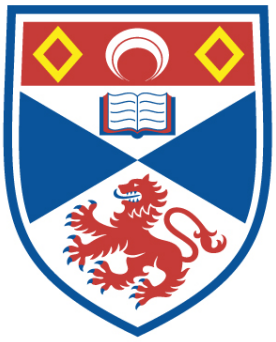
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2011

The desaturase gene family: an evolutionary study of putative speciation genes in 12 species of *Drosophila*

Maria C. Keays



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

PhD

at the
University of St Andrews

June 2011

The desaturase gene family: an evolutionary study of putative speciation genes in 12 species of *Drosophila*

Abstract

The formation and persistence of species are the subject of much debate among biologists. Many species of *Drosophila* are behaviourally isolated, meaning that heterospecific individuals are not attracted to one another and do not interbreed. Often, this behavioural isolation is at least in part due to differences in pheromonal preference. *Drosophila* pheromones are long-chain cuticular hydrocarbons (CHCs). Desaturases are enzymes that are important for the production of CHCs. This thesis investigates the evolution of the gene family across 12 species of *Drosophila*. Desaturase genes were located in all species. Some genes, those that have previously been shown to have important roles in pheromonal communication, have experienced duplication and loss in several species. Two previously undiscovered duplicates were identified. Generally the desaturase gene family is governed by purifying selection, although following duplication these contraints are relaxed and in some cases duplicated genes show compelling evidence of positive selection. One of the loci under positive selection, the novel duplicate *desat1b* of the *obscura* group, was found to have a sex-biased expression pattern and alternative splicing in its $5'$ UTR. In RNAi knock-down experiments of desaturase gene function in *D. melanogaster*, several desaturases were shown to affect CHC profiles of males and females, including some that were previously unlinked to CHC production.

Declarations

I, Maria C Keays, hereby certify that this thesis, which is approximately words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in June 2007 and as a candidate for the degree of Doctor of Philosophy in June 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2007 and 2011.

Date..... Signature of candidate.....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date..... Signature of supervisor.....

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date..... Signature of candidate.....

Date..... Signature of supervisor.....

Acknowledgements

Firstly I would like to thank Mike Ritchie and Daniel Barker for being excellent supervisors, who were always on hand to answer my questions and offer helpful feedback. The lab work would not have been possible without Tanya Sneddon, who gave me a huge amount of help and advice on everything from primer design to keeping flies happy. Jeff Graves also gave me a lot of useful suggestions when the lab work was not going so well. Anastasia Gardiner and Nadège Minois taught me about RNAi. Many thanks to Claude Wicker-Thomas who also gave me a lot of advice on this and performed all the GC-MS in her lab. Thanks to Dave Ferrier for helpful advice on cloning, and also, along with Richard Abbott, for providing useful suggestions and feedback at review meetings. Andy Goodall also helped me a lot with cloning, and Paris Veltsos helped me to analyse CHC data. Valentina Islas, Elina Immonen, Gil Smith and everyone else from the postgrad office, past and present, for keeping me sane and generally making this a fun place to be. Finally, Joe Milbourn made it possible for me to stay in St. Andrews for the last few months and is generally awesome.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Reproductive isolation and speciation

Questions regarding the formation and persistence of species have long been the subject of much debate among biologists (Wu and Ting, 2004). According to Mayr (1942), species are "groups of interbreeding natural populations that are reproductively isolated from other such groups". When reproductive isolation arises, gene flow between the populations is reduced until eventually it no longer occurs. Reproductive isolation is the result of genetic "barriers", which prevent gene flow, as opposed to physical barriers which do not allow species to come into contact with one another (Wu and Ting, 2004).

Mechanisms by which reproductive isolation acts may be classified as pre-mating or post-mating. Post-mating mechanisms include those that contribute to problems during fertilisation, or to sterility or inviability of resulting hybrid offspring (Coyne and Orr, 2004). Studies of hybrid sterility and inviability have found that the process of hybrid sterility evolves far more rapidly than that of inviability (Wu and Ting, 2004). Isolation can arise as a result of cytoplasmic incompatibilities, via chromosomal rearrangements that lead to problems during meiosis in hybrid offspring, or because of incompatibilities caused by individual "speciation genes" in the two species (Wu and Ting, 2004). Many arthropod species are infected with bacteria of the genus *Wolbachia*, which inhabit the egg cytoplasm, and have been found to cause post-zygotic reproductive isolation via cytoplasmic incompatibility

in mosquitoes, *Drosophila* and *Nasonia* (Breeuwer *et al.*, 1992; O'Neill *et al.*, 1992; Rousset *et al.*, 1992; Werren and Jaenike, 1995). Work by Noor *et al.* (2001) on *Drosophila pseudoobscura* and its sister species *Drosophila persimilis* suggests that chromosomal inversions between closely-related species create linkage groups that cause sterility in hybrid offspring, and allow species to remain separate in the face of gene flow.

Isolating mechanisms or barriers that act before mating include various ecological barriers, such as isolation via breeding season – if two species breed at different times of the year they will not mate interspecifically (Coyne and Orr, 2004). In plants, a difference in pollinator will cause pre-mating isolation because pollen will not be transferred between plants that do not share pollinators (Kephart and Theiss, 2004; Rieseberg and Wendel, 2004). Mechanical incompatibilities between individuals can also be a contributor in pre-mating isolation, for example physical incompatibility between male and female genitalia, as has been demonstrated in millipedes (Tanabe and Sota, 2008).

1.1.1 Behavioural isolation

Species can also be behaviourally isolated. This means that they do not mate heterospecifically due to a lack of attraction between the opposite sexes of different species (Coyne and Orr, 2004; Wu and Ting, 2004). Conspecific males and females recognise one another in many different ways, via visual, auditory, or chemosensory stimuli (or some combination thereof). These cues have been shown to be important in maintaining species boundaries in many different taxa. For example, several studies in butterflies have shown that differences in male wing patterning has a large effect on the mate preference of females, who prefer males of their own species (Silberglied and Taylor, 1978; Wiernasz and Kingsolver, 1992). Wiernasz and Kingsolver (1992) demonstrated that artificially colouring a heterospecific male to look like a conspecific one made them much more acceptable to the females, proving that wing patterning is largely responsible for reproductive isolation.

Auditory stimuli have also been found to contribute to behavioural isolation in a great many species. In lots of species of *Drosophila*, males "sing" to females as

part of the courtship ritual, using wing vibrations to produce sound. The songs are species-specific, and it has been demonstrated that females prefer the songs of conspecific males over those of heterospecific ones (Ritchie *et al.*, 1999). Song has been found to influence reproductive isolation in other insects, such as crickets (Hoy and Paul, 1973). It is also an important factor in many other species, such as birds (Catchpole, 1987), and frogs (Ryan and Rand, 1993).

Chemosensory stimuli in the form of pheromones are an important contributor to reproductive isolation between many species. Communication via chemical signalling between individuals has been studied in depth for many years. The term "pheromone" was coined by Karlson and Butenandt (1959), to distinguish them from hormones, which remain within an individual organism. Pheromones are substances that are secreted to the outside, and that provoke a reaction in another individual of the same species. Pheromone secretion can occur in a specific gland, or as a cuticular excretion. Their mode of reception may be olfactory or oral. Binding of the pheromone molecule to the receptor initiates a signalling cascade which ultimately leads to a behaviour in the receiving individual (Karlson and Butenandt, 1959; Wilson and Bossert, 1963). Pheromones are used by an enormous number of diverse taxa, including mammals, reptiles, insects and other arthropods, and have many different functions. Among these functions are mate recognition, territory marking, kin recognition, and attracting other members of the species to a food resource (Symonds and Elgar, 2008). The two closely-related moth species *Ostrinia nubilalis* and*O. furnicalis* use different blends of volatile hydrocarbon molecules as sex pheromones. The difference between the two species was demonstrated to be due to the activity of enzymes called desaturases, which are involved in the production of these molecules. Different types of desaturase are used during hydrocarbon synthesis, causing the resulting pheromones to differ between the two species (Roelofs *et al.*, 2002). Desaturases also have important roles in *Drosophila* sex pheromone synthesis (Chertemps *et al.*, 2006; Dallerac *et al.*, 2000; Ferveur *et al.*, 1996; Wicker-Thomas *et al.*, 1997). The action of two desaturases in *D. melanogaster*, Desat1 and DesatF, is shown in Figure 1.1. *Drosophila* pheromones are waxy, long-chain hydrocarbons which are secreted onto the fly cuticle. These non-volatile compounds form

a blend which is species-specific and in some species, such as *D. melanogaster*, is sexually dimorphic. They have been shown to be major contibutors to reproductive isolation between many species of *Drosophila* (Cobb and Jallon, 1990).

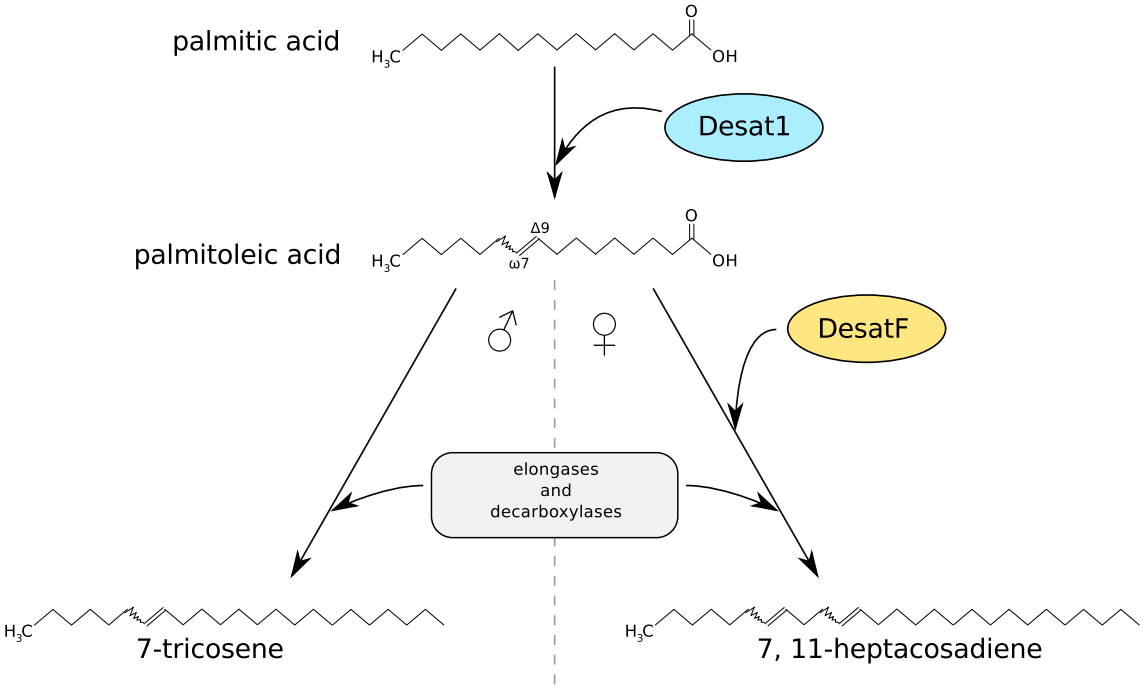


Figure 1.1: The enzymes Desat1 and DesatF are involved in the synthesis of *D. melanogaster* sex pheromones. Further explanation of the action of the desaturases is given in Chapter 2.

1.2 Gene family evolution

The desaturases form a gene family with orthologs across a great many diverse taxa, including mammals, fish, arthropods, and plants (Nakamura and Nara, 2004). Gene families grow by accumulating members via gene duplication, and shrink through gene loss. Following a duplication event, there is thought to be a period of relaxed selection on the duplicated gene, allowing it to accumulate mutations. Ultimately, the fate of duplicated genes is dependent upon the selective pressures acting on them. If the new duplicate is not required or has a negative effect on the fitness of the organism, it will become silenced by the accumulation of detrimental mutations.

This is known as pseudogenisation, or nonfunctionalisation. On the other hand, the new duplicate may acquire a novel, advantageous function as it accumulates mutations. This would lead to it being favoured by positive natural selection, or adaptive evolution, and is known as neofunctionalisation. The other copy would retain the original function. Finally, in subfunctionalisation, both copies may accumulate mutations but in different regions of their sequences, so that each copy ends up with a different function, but together they still perform the original function (Lynch and Conery, 2000; Lynch and Force, 2000; Ohta, 1994).

Studying the evolution of gene families across taxa by examining patterns of duplication and loss, and measuring selective pressures, can provide insights into their role in the evolution of species. In 2007, the complete genome sequences of 12 species of *Drosophila* (Figure 1.2) were released (*Drosophila* 12 Genomes Consortium, 2007). This has provided researchers with new opportunities for studying gene family evolution across a large amount of evolutionary divergence. Several studies of *Drosophila* gene families, including odorant binding proteins and olfactory receptors have been conducted so far (Fang *et al.*, 2009; Gardiner *et al.*, 2008; Guo and Kim, 2007; Hahn *et al.*, 2007; Vieira *et al.*, 2007). In general they have found that purifying selection is the dominant force governing gene families in these species, although these constraints are often relaxed following duplication, and in some of these cases positive selection has been detected.

1.3 Thesis objectives

The aim of this thesis is to investigate the evolution of the desaturase gene family in the 12 sequenced species of *Drosophila*. In Chapter 2, I locate orthologs of nine *D. melanogaster* desaturase loci among the species, and examine patterns of gene duplication and loss. A comparative study of *Drosophila* desaturases was recently released by Fang *et al.* (2009); my results concur with theirs, and go on to locate novel duplicates in *D. pseudoobscura*, *D. persimilis* and *D. ananassae*. In Chapter 3, I analyse the loci found in the previous chapter to measure the selective pressures acting on them, and find two cases of duplicated genes showing compelling

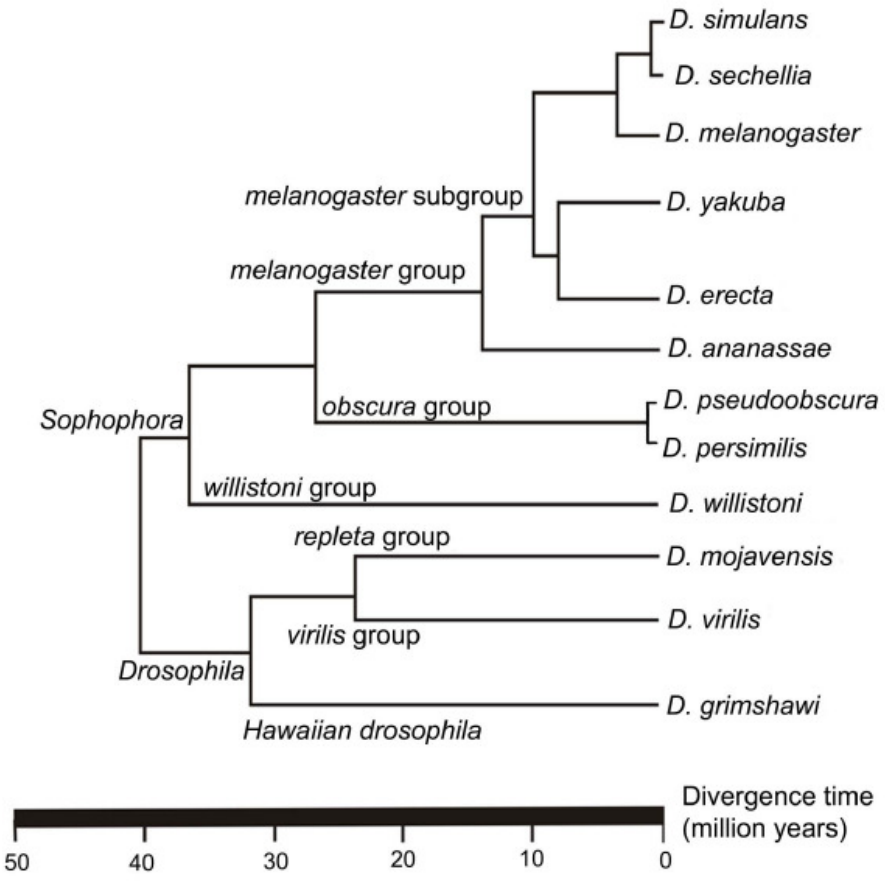


Figure 1.2: Phylogeny of the 12 sequenced *Drosophila* species. Image from `http://rana.lbl.gov/drosophila/` (*Drosophila* 12 Genomes Consortium, 2007).

evidence for positive selection. One of these is *desat1b*, a novel duplicate unique to *D. pseudoobscura* and *D. persimilis*. I attempt to further characterise this gene in Chapter 4, by analysing its expression in males and females of both species in which it was found. In Chapter 5, I use RNA interference of desaturase gene expression in *D. melanogaster*, to remove the function of several different desaturase genes, and examine the effects on the cuticlar hydrocarbon profiles of males and females.

The work in Chapters 2 to 4 is to be published in Molecular Ecology in the near future (Keays *et al.*, 2011).

Bibliography

Breeuwer, J. A. J., Stouthamer, R., Barns, S. M., Pelletier, D. A., Weisberg, W. G. and Werren, J. H. (1992). Phylogeny of cytoplasmic incompatibility microorganisms in the parasitoid wasp genus *Nasonia* (Hymerioptera: Pteromalidae) based on 16s ribosomal DNA sequences. *Insect Molecular Biology*, **1**:25–36.

Catchpole, C. K. (1987). Bird song, sexual selection and female choice. *Trends in Ecology and Evolution*, **2**:94–97.

Chertemps, T., Duportets, L., Labeur, C., Ueyama, M. and Wicker-Thomas, C. (2006). A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in *Drosophila melanogaster*. *Insect Molecular Biology*, **15**:465–473.

Cobb, M. and Jallon, J.-M. (1990). Pheromones, mate recognition and courtship stimulation in the *Drosophila melanogaster* species sub-group. *Animal Behaviour*, **39**:1058–1067.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sunderland, Massachusetts: Sinauer Associates.

Dallerac, R., Labeur, C., Jallon, J.-M., Knipple, D. C., Roelofs, W. L. and Wicker-Thomas, C. (2000). A $\Delta 9$ desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **97**:9449–9454.

Fang, S., Ting, C.-T., Lee, C.-R., Chu, K.-H., Wang, C.-C. and Tsaur, S.-C. (2009). Molecular evolution and functional diversification of fatty acid desaturases af-

ter recurrent gene duplication in *Drosophila*. *Molecular Biology and Evolution*, **26**:1447–1456.

Ferveur, J.-F., Cobb, M., Boukella, H. and Jallon, J.-M. (1996). World-wide variation in *Drosophila melanogaster* sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. *Genetica*, **97**:73–80.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C. and Ritchie, M. G. (2008). *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Molecular Ecology*, **17**:1648–1657.

Guo, S. and Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular Biology and Evolution*, **24**:1198–1207.

Hahn, M. W., Han, M. V. and Han, S. G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, **3**:e197.

Hoy, R. R. and Paul, R. C. (1973). Genetic control of song specificity in crickets. *Science*, **180**:82–83.

Karlson, P. and Butenandt, A. (1959). Pheromones (Ectohormones) In Insects. *Annual Review of Entomology*, **4**:39–58.

Keays, M. C., Barker, D., Wicker-Thomas, C. and Ritchie, M. G. (2011). Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. *Molecular Ecology*, **20**:3617–3630.

Kephart, S. and Theiss, K. (2004). Pollinator-mediated isolation in sympatric milkweeds (*Asclepias*): do floral morphology and insect behavior influence species boundaries? *New Phytologist*, **161**:265–277.

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, **290**:1151–1155.

Lynch, M. and Force, A. G. (2000). The origin of interspecific genome incompatibility via gene duplocation. *The American Naturalist*, **156**:590–605.

Mayr, E. (1942). *Systematics and the origin of species*. New York: Columbia University Press.

Nakamura, M. T. and Nara, T. Y. (2004). Structure, function and dietary regulation of $\Delta 6$, $\Delta 5$, and $\Delta 9$ desaturases. *Annual Review of Nutrition*, **24**:345–376.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. and Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the USA*, **98**:12084–12088.

Ohta, T. (1994). Further examples of evolution by gene duplication revealed though DNA sequence comparisons. *Genetics*, **138**:1331–1337.

O'Neill, S. L., Giordiano, R., Colbert, A. M. E., Karr, T. L. and Robertson, H. M. (1992). 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences of the USA*, **89**:2699–2702.

Rieseberg, L. H. and Wendel, J. (2004). Plant speciation – rise of the poor cousins. *New Phytologist*, **161**:3–8.

Ritchie, M. G., Halsey, E. J. and Gleason, J. M. (1999). *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou and Hall cycles in *D. melanogaster* song. *Animal Behaviour*, **58**:649–657.

Roelofs, W. L., Liu, W., Hao, G., Jiao, H., Rooney, A. P. and Jr, C. E. L. (2002). Evolution of moth sex pheromones via ancestral genes. *PNAS*, **99**:13621–13626.

Rousset, F., Bouchon, D., Pintureau, B., Juchault, P. and Solignac, M. (1992). *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **250**:91–98.

Ryan, M. J. and Rand, A. S. (1993). Species recognition and sexual selection as a unitary problem in animal communication. *Evolution*, **47**:647–657.

Silberglied, R. E. and Taylor, O. R. (1978). Ultraviolet reflection and its behavioral role in the courtship of the sulfur butterflies *Colias eurytheme* and *C. philodice*. *Behavioural Ecology and Sociobiology*, **3**:203–243.

Symonds, M. R. E. and Elgar, M. A. (2008). The evolution of pheromone diversity. *Trends in Ecology and Evolution*, **23**:220–228.

Tanabe, T. and Sota, T. (2008). Complex copulatory behavior and the proximate effect of genital and body size differences on mechanical reproductive isolation in the millipede genus *Parafontaria*. *The American Naturalist*, **171**:692–699.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, **450**:203–218.

Vieira, F. G., Sánchez-Garcia, A. and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology*, **8**:R235.

Werren, J. H. and Jaenike, J. (1995). *Wolbachia* and cytoplasmic incompatibility in mycophagous *Drosophila* and their relatives. *Heredity*, **75**:320–326.

Wicker-Thomas, C., Henriet, C. and Dallerac, R. (1997). Partial characterization of a fatty acid desaturase gene in *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology*, **27**:963–972.

Wiernasz, D. C. and Kingsolver, J. G. (1992). Wing melanin pattern mediates species recognition in *Pieris occidentalis*. *Animal Behaviour*, **43**:89–94.

Wilson, E. O. and Bossert, W. H. (1963). Chemical Communication among Animals. *Recent Progress in Hormone Research*, **19**:673–716.

Wu, C.-I. and Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, **5**:115–122.

Chapter 2

The Desaturase Gene Family in 12 Species of *Drosophila*

2.1 Introduction

2.1.1 Chemical communication in insects

Much of the earlier work on insect pheromonal communication was carried out on moths, due to their importance as agricultural pests (Roelofs and Bjostad, 1984). In fact the first characterised sex pheromone was that of the moth *Bombyx mori*. Named "bombykol", it is a volatile long-chain hydrocarbon released by females, and can be detected by males several kilometres away (Karlson and Butenandt, 1959; Wilson and Bossert, 1963). It is a dienic hydrocarbon, containing two unsaturations along its length. Over subsequent decades, many more Lepidopteran pheromones have been identified, and most of them are also long-chain hydrocarbons possessing between 1 and 3 unsaturations in varying positions (Roelofs and Bjostad, 1984; Roelofs and Rooney, 2003). These unsaturations are created by enzymes called desaturases.

Desaturase enzymes interact with cytochrome-b5 and cytochrome-b5-reductase in an electron transport reaction which creates an unsaturation in a hydrocarbon chain. Desaturases typically reside within a membrane and as such their secondary structure comprises four transmembrane helices. The amino acid sequence of a

desaturase also contains three highly conserved "histidine boxes", which are thought to come together in the three-dimensional structure of the enzyme and bind an iron atom, forming part of the putative active site (Nakamura and Nara, 2004). More recently, studies have found that the pheromones of *Drosophila* also take the form of long-chain hydrocarbons, and that desaturase enzymes also play a pivotal role in their biosynthesis.

2.1.2 *Drosophila* courtship and cuticular hydrocarbons

Drosophila courtship involves a number of different behaviours which are mediated by various cues. A male may first be attracted to a female by a visual cue, which leads to "orientation" behaviour: the male begins to follow the female. The male will try to touch the female with a foreleg, detecting phermones on the cuticle of the female's abdomen. If the pheromones are attractive to the male he will progress further in courtship by "singing" – vibrating one of his wings to produce sound – and by licking the female's genitalia with his proboscis (Greenspan and Ferveur, 2000; Hall, 1994). The female is also able to detect the male's pheromones, and is thought to base her decision on whether to mate on a combination of her preference (or lack thereof) for his song and his pheromones (Ferveur *et al.*, 1997; Grillet *et al.*, 2006). *Drosophila* pheromones take the form of long-chain fatty acid hydrocarbons that are secreted onto the fly cuticle by cells called oenocytes (Ferveur, 2005; Ferveur *et al.*, 1997; Jallon, 1984), and are known as cuticular hyrocarbons (CHCs). As well as acting as sex pheromones, they are thought to be important in insect physiology, playing roles in traits such as cold tolerance and resistance to desiccation (Rouault *et al.*, 2000). CHCs may be saturated – with no double bonds – or they may be unsaturated, containing one or two double bonds (unsaturations) at specific positions. Mono-unsaturated CHCs are called monoenes, and di-unsaturated ones are called dienes. Fly pheromones are blends of many different CHCs, which are species-specific. Certain species of the *Drosophila melanogaster* subgroup (for example *D. melanogaster*, *D. sechellia* and *D. erecta*), are sexually dimorphic with respect to their CHCs. The CHCs of males and females have different numbers of unsaturations: males of the *melanogaster* subgroup tend to have high levels of

monoenes – meaning they contain one unsaturation – while females tend to have a CHC profile rich in dienes (Chertemps *et al.*, 2006; Cobb and Jallon, 1990; Dallerac *et al.*, 2000). This is not true for all species however, for example *D. simulans*, *D. mauritiana* and *D. yakuba*, which are sexually monomorphic. In these species, the males and females share the same, monoenoic, CHCs (Chertemps *et al.*, 2006; Cobb and Jallon, 1990; Jallon and David, 1987). The main CHCs of some species in the *D. melanogaster* species subgroup are summarised in Table 2.1. CHCs have been shown to be important in the courtship of species outside the *D. melanogaster* subgroup, including *D. ananassae* (Doi *et al.*, 1997), *D. virilis* (Liimatainen and Jallon, 2007) and its close relatives (Liimatainen and Hoikkala, 1998), *D. mojavensis* (Etges and Jackson, 2001) and *D. grimshawi* (Droney and Hock, 1998).

			Main CHC	
	Species	Strain	Female	Male
Dimorphic	*D. erecta*		9,23 TTCD	7 T
	D. sechellia		7,11 HD	6 T
	D. melanogaster	Cosmopolitan	7,11 HD	7 T
		Some African/Caribbean	5,9 HD	7 P
Monomorphic	*D. simulans*	Central Africa	7 P	7 P
		Elsewhere	7 T	7 T
	D. mauritiana		7 T	7 T
	D. yakuba		7 T	7 T

Table 2.1: The main CHCs of some species in the *D. melanogaster* subgroup. TTCD: Tritricontadiene; HD: heptacosadiene; T: tricosene; P: pentacosene. After Cobb and Jallon (1990).

Cobb and Jallon (1990) found that males would court females from species other than their own, or sometimes even other males, apparently based on their CHC profile. For example, *D. melanogaster* and *D. sechellia* females both exhibit high levels of 7,11-heptacosadiene (7,11-HD). Cobb and Jallon (1990) found that *D. melanogaster* males would court *D. sechellia* females almost as much as, or in some cases more than, they would court conspecific females. They also observed that males from monomorphic species would court females from other monomorphic species, but would largely ignore females from dimorphic species. A small amount of male-male courtship was observed, although interestingly the males showing the highest amount of courtship towards other males were *D. melanogaster* ones – perhaps surprisingly because their conspecific females have CHC profiles high in dienes, and none of the male flies possess dienes. Also, *D. melanogaster* males court *D. erecta* and *D. sechellia* males to roughly the same extent, though *D. erecta* males possess 7-tricosene (7-T) and *D. sechellia* males have 6-T. Cobb and Jallon (1990) therefore conclude that while some of their data provides evidence that CHCs are important in mate recognition and courtship activity, many other factors, such as visual and auditory perception for example, must be involved in inducing courtship behaviour and in determining its outcome (Cobb and Jallon, 1990).

The evidence that CHC profiles are important in reproductive isolation was later reinforced by Coyne *et al.* (1994), who showed that the courtship behaviour of *D. simulans* males is strongly influenced by the type of CHC possessed by the females, regardless of their species. As noted by Cobb and Jallon (1990), males of *D. simulans* only very rarely court females of *D. sechellia*, who possess dienes. In their experiment, Coyne *et al.* (1994) managed to transfer CHCs between females of *D. simulans* and *D. sechellia* by crowding the flies together in vials. They found that this crowding caused transfer of CHCs so that one female would acquire almost half the amount of the predominant CHC of the other species. They demonstrated that *D. simulans* males largely ignored conspecific females that had undergone the transfer of CHCs with *D. sechellia* females and thus carried 7,11-HD, though at the same time they would readily court ordinary *D. simulans* females. They note that this pattern remained consistent even when the females were killed, meaning

that the lack of courtship was not likely to have been due to some effect on female behaviour caused by crowding. They also report that the opposite experiment, with *D. sechellia* females who had acquired 7-T from the *D. simulans* females, in turn showed promising results: the *D. sechellia* females carrying the 7-T induced more courtship from *D. simulans* males than did the ordinary *D. sechellia* females. This increase in attraction is not as pronounced as the decrease observed in the first part of the experiment. They also carried out the same experiments using *D. mauritiana* and *D. melanogaster* flies as well as *D. simulans* ones, and found that crowding of *D. simulans* and *D. mauritiana* females (who also possess 7-T) made no difference to the attractiveness of the *D. simulans* females to conspecific males; however crowding *D. simulans* females with *D. melanogaster* females (with 7,11-HD) caused a marked reduction in courtship behaviour from the *D. simulans* males (Coyne *et al.*, 1994). Similar "perfuming" experiments have been carried out in other species. Blows and Alan (1998) performed a similar analysis on the closely-related species *D. serrata* and *D. birchii*, and showed that CHC profile has a major influence on mate choice in these two species; Etges and Ahrens (2001) used a perfuming experiment to demonstrate that CHC profiles are important in the reproductive isolation between different populations of *D. mojavensis*.

Further insight into the role of CHCs in courtship and reproductive isolation has been gained via the study of transgenic *D. melanogaster* lines. Savarit *et al.* (1999) produced lines in which they were able to induce, via heat-shock, over-expression of the *transformer* (*tra*) gene, which is important in sex determination (Butler *et al.*, 1987). In females, this over-expression caused almost all CHCs, including all known pheromonal compounds, to be completely eliminated. They performed mating experiments using these females, with the expectation that they would be unattractive to conspecific males. However, not only were they attractive to *D. melanogaster* males, they were also actively courted by males of *D. simulans*, *D. sechellia* and *D. mauritiana*. They found that while most CHCs are eliminated by the over-expression of *tra*, five separate hydrocarbons are still produced. They suggest that one or more of these could be an ancestral form of attractant, and that known pheromones such as 7,11-HD reinforce this attraction in conspecifics, and repel het-

erospecifics. Billeter *et al.* (2009) later managed to generate *D. melanogaster* flies which lack oenocytes, and therefore cannot produce any CHCs. Their experiments showed that oenocyte-lacking (oe^-) females were actively courted by conspecific wild type males, who in fact preferred them to wild type females. oe^- males were also courted by wild type *D. melanogaster* males. oe^- females were attractive to heterospecific males: *D. simulans*, *D. yakuba* and *D. erecta* males all courted them. The importance of 7,11-HD as a reproductive isolation barrier was demonstrated by coating oe^- females with this compound and repeating these mating experiments – this time they were unattractive to heterospecific males. In another experiment, they treated oe^- females with different amounts of 7,11-HD and *cis*-vaccenyl acetate (cVA), the male aversive pheromone. They found that while cVA delayed mating, 7,11-HD appeared to counteract its effect. They suggest that this could be a case of sexual conflict, whereby 7,11-HD solicits further matings despite the cVA left behind by the previous male. They hypothesise that, given both male and female oe^- flies are sexually attractive to wild type males, CHCs act to confer sexual identity onto an otherwise attractive fly substrate. They show that in *D. melanogaster* 7,11-HD has multiple functions – reinforcing attraction in conspecifics, and repelling heterospecifics.

2.1.3 The *desat1* locus

The enzyme Desat1 is known to be responsible for the introduction of the double-bond in monoenes, and for the first double-bond in dienes, at least in *D. melanogaster* and its close relatives (Figure 1.1). It was first isolated by Wicker-Thomas *et al.* (1997), and its amino acid sequence found to contain the highly conserved histidine-rich regions essential for desaturase action. Later, Dallerac *et al.* (2000) showed that it is able to create unsaturated fatty acid precursors of $\omega 7$ hydrocarbons, like those used in CHC biosynthesis.

Marcillac *et al.* (2005a) generated mutant *D. melanogaster* flies with a *PGal4* transposon inserted into the regulatory region of *desat1*. They show that this mutation disrupted the expression of *desat1* in homozygous mutants, with the result that they produced between 70 and 90% fewer sex pheromones than control flies.

They also found that sex-specific differences were almost completely removed in the mutant flies. They were able to rescue the pheromonal phenotypes of mutants by creating alleles with all or most of the inserted transposon removed, showing that the presence and size of the insertion affected the expression and function of *desat1*. In the same study, Marcillac *et al.* (2005a) also detected five individual splice isoforms of the *desat1* gene in control flies, each with a different non-coding exon at the 5′ end. Three minor transcripts showed variation between males and females, while the most prevalent transcript showed no discernible sexual dimorphism. In the mutant flies, levels of all transcripts were greatly reduced. They also analysed the *desat1* transcripts of *D. simulans*, which does not show sexually dimorphic CHCs, and report that the two transcripts exhibiting sexual dimorphism in *D. melanogaster* were not found in *D. simulans*. This indicates that these *desat1* transcripts have a role in influencing sex-specific CHCs in *D. melanogaster*.

Work from the same research group also indicates that as well as being responsible for creating sex pheromones, *desat1* has a role in their perception. (Marcillac *et al.*, 2005b). The same mutant flies were used, with the transposon inserted in the *desat1* regulatory region. They show that not only is the production of unsaturated CHCs greatly reduced by this, making mutant males and females indistinguishable to control males, but also that mutant males were unable to distinguish between control males and females in mating experiments. They find that *desat1* is expressed in tissues involved in detecting sex pheromones, as well as in the oenocytes, which produce CHCs. Both mutant phenotypes – the reduction in unsaturated CHCs and the disrupted perception – were rescued by excising the transposon. These intriguing results indicate that the regulatory region of *desat1* may have been shaped by evolution to produce pleitropic activity (Bousquet *et al.*, 2009; Marcillac *et al.*, 2005b). Further work on this phenomenon by Houot *et al.* (2010) has shown that distinct *desat1* regulatory regions and/or transcripts are separately controlling the production and perception of CHCs in *D. melanogaster*.

2.1.4 The *desatF* locus

The formation of the second double-bond of dienic CHCs, as found in *D. melanogaster* and *D. sechellia* females, requires a second desaturation step which is performed by a second desaturase enzyme (Figure 1.1; Gleason *et al.*, 2005). Previous studies indicated that the gene responsible for the sexual dimorphism was likely to reside on the third chromosome in *D. melanogaster*, in region 67E–69B (Coyne, 1996; Gleason *et al.*, 2005). Chertemps *et al.* (2006) have recently characterised a gene in this region that is probably responsible for the dimorphism in *D. melanogaster*. They name it *desatF*. Using Northern blot analysis, they found that it is only expressed in females of *D. melanogaster* and *D. sechellia*, which have high levels of dienes. They detected no expression in *D. simulans* females or males, despite the gene being present in the genome and the probe hybridising strongly to *D. simulans* genomic DNA although it was based on the *D. melanogaster* sequence.

To discover whether *desatF* was likely to have an effect on CHC biosynthesis, Chertemps *et al.* (2006) used RNA interference (RNAi) knock-down to try and block the expression of the gene, and also overexpression of *desatF* using the GAL4/UAS system, which enables targeted expression of genes of interest (Duffy, 2002). They found that RNAi knockdown of *desatF* in females caused a dramatic drop in the amounts of 7,11-HD (−83%) and another diene, 7,11-nonacosadiene (7,11-ND) (−85%). This was accompanied by a large increase in the amounts of monoenes. Overall, they note that the increase in monoenes (550ng) was almost equal to the decrease in dienes (594ng). This supports the idea that *desatF* is responsible for production of the second double bond in dienes. The overexpression analysis points to a similar conclusion: it produced an increase in 7,11-HD of 57%, and a decrease of 50% in 7-T.

Chertemps *et al.* (2006) also tested the effect of the RNAi knockdown and overexpression of *desatF* in females on the courtship behaviour of males. They found that females with overexpressed *desatF* induced 60% more copulation attempts than control females, while the RNAi knock-down treatment caused a decrease in the number of copulation attempts and the amount of courtship, accompanied by a higher courtship latency. They also induced feminisation of male CHCs by express-

ing the *transformer* gene. They detected a large amount of *desatF* expression in males that had undergone the feminisation, whereas there was none in control males. The overexpression and RNAi knock-down treatments in males did not produce any differences in the CHC profile. This is not surprising given that no *desatF* expression was detected in males. Chertemps *et al.* (2006) propose that the lack of effect of overexpression suggests that either the substrate required by *desatF* is not available in males, or that the product of *desatF* cannot be elongated to make the final pheromone product. This in turn suggests that there must be other female-specific genes involved that have not yet been characterised (Chertemps *et al.*, 2006).

All the results from the experiments by Chertemps *et al.* (2006) indicate that *desatF* has a crucial role in pheromone biosynthesis in *D. melanogaster* females. The fact that it only appears to be expressed in females that have dienes as their predominant CHCs suggests that its action contributes to behavioural isolation between sexually dimorphic and non-dimorphic species, and therefore its evolution may have been important in speciation (Chertemps *et al.*, 2006).

Further work on *desatF* has recently been carried out by Legendre *et al.* (2008). In the experiments of Chertemps *et al.* (2006), although RNAi knock-down greatly decreased the levels of 7,11-HD, it was not completely absent. This could have been caused by a small amount of *desatF* expression if the RNAi did not block all *desatF* transcripts, or perhaps due to the action of other genes which have not been characterised (Chertemps *et al.*, 2006). To try to solve this issue, Legendre *et al.* (2008) crossed *D. simulans* males and *D. melanogaster* females to create hybrid females – some that were wild-type, and therefore had one working copy of *desatF* from *D. melanogaster*, and some that carried a deletion in the region containing *desatF* and therefore could not express the gene. They found that the hybrids with the functional copy of *desatF* produced about half the level of dienes found in *D. melanogaster*, but that a high proportion of them were shorter than ordinary *D. melanogaster* ones. This indicates that perhaps the enzymes involved in elongation of the hydrocarbons in diene-carrying females are sensitive to gene dosage (Legendre *et al.*, 2008). The hybrids with the deletion in the *desatF* region did not express any dienes at all, therefore Legendre *et al.* (2008) concludes that *desatF* must be

the only gene responsible for their production.

Chertemps *et al.* (2006) note that *desatF* is expressed in females of both 7,11-HD- and 5,9-HD-carrying strains of *D. melanogaster*. Legendre *et al.* (2008) used RNAi knock-down of *desatF* in females of a hybrid between the Tai and Canton-S strains, to determine whether levels of 5,9-HD are affected and thus whether *desatF* is also likely to be important for 5,9-HD production as well as that of 7,11-HD. The resulting RNAi knock-down hybrid showed strong inhibition of both types of dienes, accompanied by a large increase in 7- and 5-monoenes. This indicates that *desatF* is able to act on both type of monoenes to produce each type of diene in either strain of *D. melanogaster*.

Legendre *et al.* (2008) tested the effect of the different CHC profiles on courtship behaviour. They found that *D. melanogaster* males actually showed reduced courtship activity towards the hybrid females carrying the functional copy of *desatF* and therefore displaying dienes. They explain that this unexpected result may be due to the fact that the dienes exhibited by these females are shorter than usual, and that this is thought to have a detrimental effect on male courtship behaviour. The dienes almost completely inhibited courtship behaviour in males of *D. simulans*, while the females carrying the deletion, and therefore lacking in dienes, induced patterns of behaviour similar to that observed towards *D. simulans* females (Legendre *et al.*, 2008).

Finally, Legendre *et al.* (2008) examined the DNA sequence variation at the *desatF* locus between *D. melanogaster* and *D. simulans*, to determine whether the promoter region of the *D. simulans* copy has mutations that would explain its lack of expression in this species. They found many short insertions in the region $5'$ to the putative start codon in *D. simulans*, but apparently none that would interrupt any of the transcription factor binding sites found. When comparing the coding regions of the locus, they found that the codon usage bias and GC content were conserved between the two species, indicating that the silencing of *D. simulans desatF* is likely to be recent, and has not yet had much time to accumulate mutations. Overall, Legendre *et al.* (2008) shows that this locus differs by roughly 25% between the two species. Given that, on average, the two species differ from one another by between

four and eight percent, a 25% difference is relatively large. Therefore, Legendre *et al.* (2008) proposes that *desatF* has undergone rapid evolution in *D. simulans* due to its silencing. This hypothesis has yet to be proven with further analysis of the region.

Shirangi *et al.* (2009) examined the evolution of *desatF* expression in seven species of *Drosophila*. They show, like previous studies, that it is expressed female-specifically in *D. melanogaster*, in the oenocytes. They also examined the expression in six other species, and find that its expression corresponds with diene production in all species. It is female-specific in *D. melanogaster* and *D. sechellia*, and not expressed in either sex of *D. simulans* or *D. yakuba*. It was found to be expressed in both males and females of *D. takahashii*, *D. serrata* and *D. pseudoobscura*, in which dienes are produced by both sexes. They show that the *desatF* gene has undergone rapid evolution across the genus, documenting six separate instances of gene loss, two modifications in sex-specific expression, and three separate instances of silencing without gene loss. They go on to discover that the rapid changes in *desatF* expression pattern is due to gains and losses of a binding site for the Doublesex transcription factor (Shirangi *et al.*, 2009). The *doublesex* gene produces male- and female-specific transcription factors which are responsible for regulating sex-specific expression of many genes (Erdman *et al.*, 1996). Shirangi *et al.* (2009) show that *desatF* is under the control of Doublesex in *D. melanogaster* females, and that the transitions in *desatF* expression from sexually monomorphic to dimorphic (and vice versa), are directly affected by gains and losses of Doublesex binding sites in the regulatory region of the gene. They suggest that these changes in *desatF* expression reflect changing regimes of sexual selection, and that *desatF* is a promising candidate for a speciation gene affecting premating isolation in *Drosophila*.

2.1.5 The *desat2* locus

The main cuticular hydrocarbon (CHC) in *D. melanogaster* females exhibits geographic variation, with females from west Africa and the Caribbean (known as "Z" type) possessing one form, and females from the rest of the world ("Cosmopolitan") possessing the other. The two forms differ in the positions of the double-bonds found

along the hydrocarbon chain: the Z females produce high levels of 5,9-heptacosadiene (5,9-HD), whereas the Cosmopolitan ones produce 7,11-HD (Ferveur *et al.*, 1996).

The gene thought to be responsible for this variation, *desat2*, was discovered by Dallerac *et al.* (2000). They found it while looking for different desaturase isoforms in two strains of *D. melanogaster*, to try to explain the observed difference in CHC profiles (Dallerac *et al.*, 2000). The strains they tested were Tai, a Z-type strain from the Ivory Coast, and Canton-S, a Cosmopolitan-type strain from Ohio. They found that this gene is expressed in females of the Tai strain only, and were unable to detect a promoter region in the Canton-S strain. These results, along with the fact that Coyne *et al.* (1999) had previously mapped the polymorphism to the same chromosomal region, led them to suggest that its activity results in the production of 5,9-HD in females of the Z type (Coyne *et al.*, 1999; Dallerac *et al.*, 2000). Their conclusion was reinforced soon after by Takahashi *et al.* (2001), who discovered a 16bp deletion about 150bp 5′ to the translation start codon (the putative promoter region) which shows a complete association with the high-5,9-HD phenotype.

It was initially assumed that, due to this deletion, the Cosmopolitan allele is inactive, and that this inactivity allows for the production of 7,11-HD (Dallerac *et al.*, 2000; Takahashi *et al.*, 2001). This was corroborated by the discovery by Takahashi *et al.* (2001) that one African line with the 7,11-HD phenotype had a deletion and a stop codon within the coding sequence of the *desat2* gene. However, Coyne and Elwyn (2006a) found that females carrying one Cosmopolitan-type allele, hemizygous with a deficiency at the *desat2* locus, produced less 7,11-HD than females homozygous for the Cosmopolitan allele. This suggests that this allele may not be completely non-functional (Coyne and Elwyn, 2006a). This finding is supported by work by Michalak *et al.* (2007), who demonstrated via microarray analysis that the gene is still transcribed in Cosmopolitan females, although it is down-regulated. The CHC profile of the male flies is not affected by the polymorphism (Dallerac *et al.*, 2000).

Takahashi *et al.* (2001) also tested whether the worldwide spread of the Cosmopolitan allele, which contains the deletion, was likely to have been driven by natural selection. To do this, they used hitch-hiking analysis of the region surrounding

the *desat2* promoter to demonstrate that the African/Caribbean populations are 2.5–3 times as variable as the Cosmopolitan populations, in the region near to the deletion. Thus, they conclude that the Cosmopolitan allele has been driven to high frequency throughout the world by positive natural selection (Takahashi *et al.*, 2001).

The two races of *D. melanogaster* sometimes show asymmetric sexual isolation when they exist in sympatry: Z-type females and Cosmopolitan males do not mate readily with one another, whereas Cosmopolitan females and Z males do (Wu *et al.*, 1995). Coupled with the knowledge of the polymorphic CHC profile, it seems logical that the observed sexual isolation is due to Cosmopolitan males failing to court the Z females based on their lack of 7,11-HD due to their *desat2* genotype. However, this is probably not the case, for several reasons. One reason is that in Caribbean populations whose females have the 5,9-HD phenotype, the pattern of sexual isolation is not seen (Takahashi *et al.*, 2001). Secondly, it is likely that the phenotypic difference is the result of more than one gene: Ting *et al.* (2001) showed, via construction of recombinant lines carrying different lengths of Z-type and Cosmopolitan-type segments of the third chromosome, that there are at least four loci from this chromosome contributing in some way to the overall Z-type behaviour (Takahashi and Ting, 2004; Ting *et al.*, 2001). Also, Greenberg *et al.* (2003) showed that reintroduction of the Z-type allele to an otherwise Cosmopolitan genome restores only about half of the level of 5,9-HD seen in Z lines. This indicates that although *desat2* seems to be the main gene affecting this phenotype, there are likely to be other genes contributing in the Z females (Greenberg *et al.*, 2003). Thirdly, and perhaps most importantly, studies of mating behaviour have shown that it is the Z-type females who show the discriminatory behaviour, and not the Cosmopolitan-type males. Coyne and Elwyn (2006a) found that, during no-choice tests where a Z-type female and a Cosmopolitan-type male were observed at intervals over a three-hour period, the courtship activity of the male remained constant. They therefore conclude that any sexual isolation between the two strains must be mainly down to discrimination by the females, not the males (Coyne and Elwyn, 2006a). Grillet *et al.* (2006) showed that females discriminate against males on the basis of their CHCs, so it is likely

that the Z-type females are repelled by the CHCs of the Cosmopolitan males.

It seems unlikely *desat2*, a gene which affects the CHC profile of females and not males, could also somehow affect female selectivity in the Z populations and at the same time produce a male phenotype that is preferred by Z females (Coyne and Elwyn, 2006a). However, Greenberg *et al.* (2003) proposed that *desat2* causes this reproductive isolation as a by-product of selection for a stress-resistance phenotype. They created transgenic lines possessing either the Z-type *desat2* gene ($ds2^Z$) or the Cosmopolitan-type one ($ds2^M$), and subjected the flies to environmental stress such as cold temperatures, starvation, and desiccation. They reported strong evidence that both males and females of the $ds2^Z$-carrying flies were less cold-resistant and more starvation resistant than the $ds2^M$-carrying ones. They did not find any difference in the desiccation resistance of either line (Greenberg *et al.*, 2003). Greenberg *et al.* (2003) therefore conclude that *desat2* is likely to be involved in ecological adaptation as well as sexual isolation. This conclusion was later challenged by Coyne and Elwyn (2006a), who report that they were unable to replicate these results. In fact, they report that in two of the cold tolerance tests the result was statistically significant in the opposite direction to that expected, with Z flies surviving better than Cosmopolitan ones. They also report inconsistencies in the starvation tests, again finding some results opposite to those expected (Coyne and Elwyn, 2006a). They propose that the discrepancy between their findings and those of Greenberg *et al.* (2003) may be attributable to the likely quantitative nature of stress tolerance traits, which can be affected by many loci that vary between the different strains used in the construction of the transgenic lines. They therefore conclude that *desat2* may not play any role in stress resistance, and the differences reported by Greenberg *et al.* (2003) and Coyne and Elwyn (2006a) are simply down to the differences in the genetic background, or that it plays a very minor role that cannot be discerned in these experiments due to the effects of variation in the genetic background (Coyne and Elwyn, 2006b). The adaptive significance of the different *desat2* alleles is still being debated. It remains to be seen whether either allele conveys some kind of ecological advantage in its respective environment, and to what extent it contributes to sexual isolation between Z-type females and Cosmopolitan-type males (Coyne and

Elwyn, 2006a).

Michalak *et al.* (2007) carried out microarray analyses of Cosmopolitan and Z females, to determine which genes show differential expression between the two races. They also compared microarray expression data from mated and non-mated Z females, to see if mating experience affected the gene expression pattern. This may help to shed light on what it is that causes the discriminatory behaviour of these females towards the Cosmopolitan males (Michalak *et al.*, 2007). They discovered 45 genes that had simultaneously different expression profiles between Z-type and Cosmopolitan females *and* between mated and unmated Z females. For *desat2*, they were able to confirm that it is over-expressed in Z females relative to Cosmopolitan ones – and as mentioned above they found that contrary to what was previously thought, *desat2 is* expressed in Cosmopolitan females despite its deletion, though it is down-regulated. However, they did not detect a significant difference in the expression pattern of *desat2* between mated and unmated Z-type females (Michalak *et al.*, 2007). This indicates that if *desat2* is involved somehow in the sexual isolation, its action is probably not affected by mating experience. Of the remaining candidate genes identified by Michalak *et al.* (2007), one was *Odorant receptor 63a* (*Or63a*), which plays an important role in olfaction and mate recognition in *Drosophila*. In their analyses, *Or63a* was downregulated in Z-type females compared with Cosmopolitan-type ones, and simultaneously suppressed in mated Z-type females relative to nonmated ones. This shows the pattern they were expecting – a Z-type characteristic appears to be "reinforced" after mating. This gene, along with the others on their list, could be a potential candidate for explaining the sexual isolation between Z-type and Cosmopolitan populations (Michalak *et al.*, 2007).

2.1.6 Twelve *Drosophila* genomes

In 2007, the complete genome sequences of 12 *Drosophila* species became publicly available (*Drosophila* 12 Genomes Consortium, 2007). This has provided evolutionary biologists with a new opportunity to examine the evolution of gene families across this genus, spanning a large amount of divergence.

In this chapter, comparative genomics techniques are employed in an attempt

to discover orthologs of the well-studied, and less well-studied, *D. melanogaster* desaturase loci among the 12 species. A similar study was published in 2009 by Fang *et al.*, in which the authors used a combination of homology searching and synteny analysis to identify desaturase orthologs among the same 12 species. The work presented in this chapter concurs with many of the findings of Fang *et al.* (2009), and goes on to reveal previously undiscovered desaturase loci in three species. A summary of the work presented in this chapter appears in Keays *et al.* (2011).

2.2 Methods

2.2.1 *D. melanogaster* desaturases

An "All Text" search among the *D. melanogaster* loci in FlyBase (Wilson *et al.*, 2008) for "desaturase" currently returns 13 results. Of these, four are well-studied and known not to code for enzymes involved in CHC manufacture: *infertile crescent* (*ifc*), *schnurri* (*shn*), *tra* and *white* (*w*). Of the remaining nine, six are annotated as having "stearoyl-CoA 9-desaturase activity": *desat1*, *desat2*, *desatF*, *CG8630*, *CG15531*, *CG9743*. Another, *CG9747*, is anotated with "acyl-CoA Δ11-desaturase activity", and a further two loci (*CG17928* and *Cyt-b5-r*) are annotated simply with "fatty acid desaturase activity". These nine loci were then used as query sequences in BLAST searches against the 12 genomes.

2.2.2 Reciprocal BLAST

All sequences were downloaded from FlyBase: the protein sequences of the nine *D. melanogaster* loci, the 12 whole genome sequences, and files containing predicted genes and their translations (*Drosophila* 12 Genomes Consortium, 2007). All BLAST searches were performed using a local installation of `blastall` version 2.2.21 (Altschul *et al.*, 1990). BLAST databases were created with `formatdb` using the whole genome Fasta files from FlyBase.

Each of the nine *D. melanogaster* protein sequences was used as the query in a TBLASTN search against each of the 12 genomes. The TBLASTN program

takes a protein sequence as the query against a DNA database. The DNA sequences are translated in all six possible reading frames, and the translations are compared with the query sequence. This method is more sensitive than a nucleotide-nucleotide BLAST, which, due to the degeneracy of the genetic code, might miss more distantly-related sequences which had divergent nucleotide sequences but more similar amino acid sequences. Thus, TBLASTN enables the detection of both closely- and distantly-related loci. The alignments in each TBLASTN output file was assessed manually to obtain the positions of potential start and end codons, and the corresponding sequence was copied from the whole genome sequence using Perl. The next step is to locate the exon-intron boundaries, so that the coding sequence could be determined.

GeneWise version 2.2.0 (Birney *et al.*, 2004) was used to align each nucleotide sequence obtained in the first round of TBLASTN to the *D. melanogaster* amino acid sequence used as the query. The GeneWise output shows the predicted positions of the intron-exon boundaries in the nucleotide sequence. These positions were used to create the coding sequence, which was in turn translated to give the putative amino acid sequence. The GeneWise and translation stages also enabled identification of any premature stop codons, which would indicate pseudogenisation.

The amino acid sequence was then used to do a reciprocal TBLASTN search: back against the *D. melanogaster* nucleotide database. If the top hit in this search was the locus used as the query in the initial round of TBLASTN, then the new locus was taken to be an ortholog of the *D. melanogaster* locus.

2.2.3 Further ortholog detection

MEME and MAST (Bailey *et al.*, 2006) were used to detect any orthologs that may have been missed in the reciprocal BLAST. MEME is a tool for detecting shared, ungapped patterns in a set of nucleotide or amino acid sequences, and produces HTML output. Its sister program, MAST, reads this output and searches for the presence of the pattern detected by MEME in a second set of sequences. All amino acid sequences discovered thus far were scanned by MEME. MEME detected the pattern show in Figure 2.1, which includes one of the histidine boxes known to

be functionally important and highly conserved in desaturases. MAST was used to search for the resulting pattern output among the predicted proteins from each species, generated by *Drosophila* 12 Genomes Consortium (2007). Novel hits in the MAST output (i.e. any not already found using BLAST) were discovered using Perl.

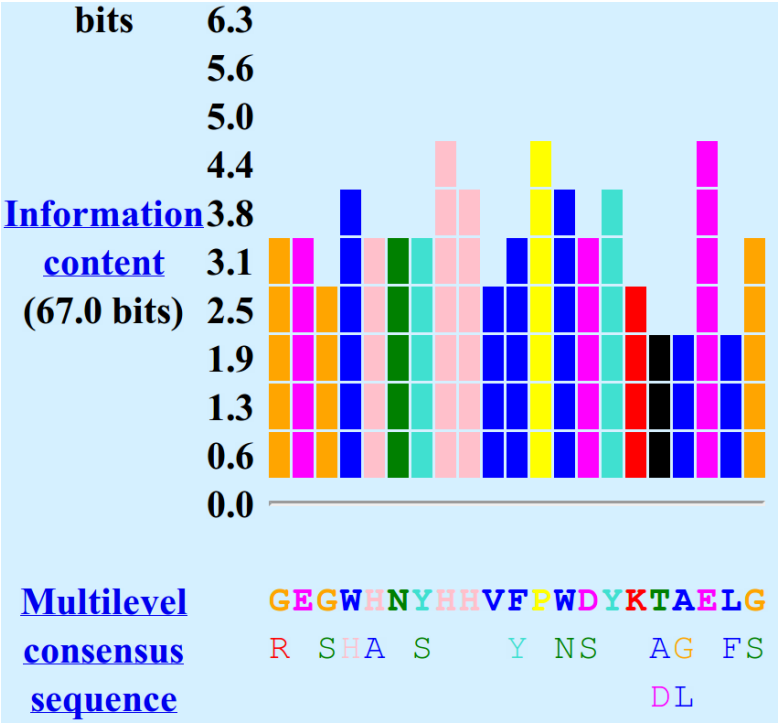


Figure 2.1: Consensus motif detected by MEME. The higher the "Information content", the more highly conserved the residue. MAST accepts the output of MEME as input and searches for the presence of the motif in a set of sequences (Bailey *et al.*, 2006).

Finally, all loci found were used in an "all-against-all" reciprocal TBLASTN search against all genomes to confirm orthology and pick up any further undetected loci.

2.2.4 Re-sequencing

Two genes for which the database sequence was incomplete or questionable, were re-sequenced. In the *D. sechellia* genome sequence file, the contig "scaffold_0" contains stretches of 'N' characters. One of these stretches coincides with the 3' end of

Target	Forward primer	Reverse primer
D. sechellia desat1	GGAGAACATCTCGGTGGCTA	TGGGCTTCTGCTCTGTTAGG
D. yakuba desatF	GAAGTACTACCTGCTCCTGATGC	GCACGCTGTTTCACCAAAT

Table 2.2: Primers used to amplify the regions to be sequenced

the putative *desat1* gene. The sequence of *D. yakuba desatF* has a 10bp deletion which results in a frameshift and premature stop codons. In FlyBase, the gene is annotated as having a 32bp intron, which corrects the frameshift. However, this intron is unlikely to be real: if it is electronically spliced out and the spliced coding sequence translated, the resulting amino acid sequence lacks a region which is highly conserved in all other DesatF proteins (Figure 2.2). The fact that this region is highly conserved suggests that it is important for the enzyme to function. Whether or not the intron is real, if the deletion is indeed present, this copy of *desatF* would not be able to produce a functional desaturase enzyme in *D. yakuba*.

Two strains of *D. yakuba* were sequenced: 14021-0261.00, and the strain used in the whole genome project, 14021-0261.01. One strain of *D. sechellia* was sequenced. DNA was extracted using the single fly DNA prep protocol (Appendix A, Section A.1). Primer3 (Rozen and Skaletsky, 2000) was used to design primers to target the 3' end of *D. sechellia desat1* and the region of *D. yakuba desatF* which contains the deletion. The primer sequences are shown in Table 2.2. The target regions were amplified using the PCR protocol shown in Appendix A, section A.2. Products were visualised on a 2% agarose gel. The products were purified using the QIAquick Gel Extraction Kit (QIAGEN; Cat. No. 28704). The sequencing reactions were performed using both forward and reverse PCR primers by DNA Sequencing & Services at the University of Dundee, who provided chromatographs and the sequences in FASTA format.

2.2.5 Phylogenetics

Phylogenies were reconstructed using the amino acid sequences, and corresponding nucleotide sequences, of all loci found. The *Saccharomyces cerevisiae* sequence OLE1 (YGL055W) was used as an outgroup. This sequence appears to be the only $\Delta 9$ desaturase gene present in *S. cerevisiae*, and was found using the Ensembl

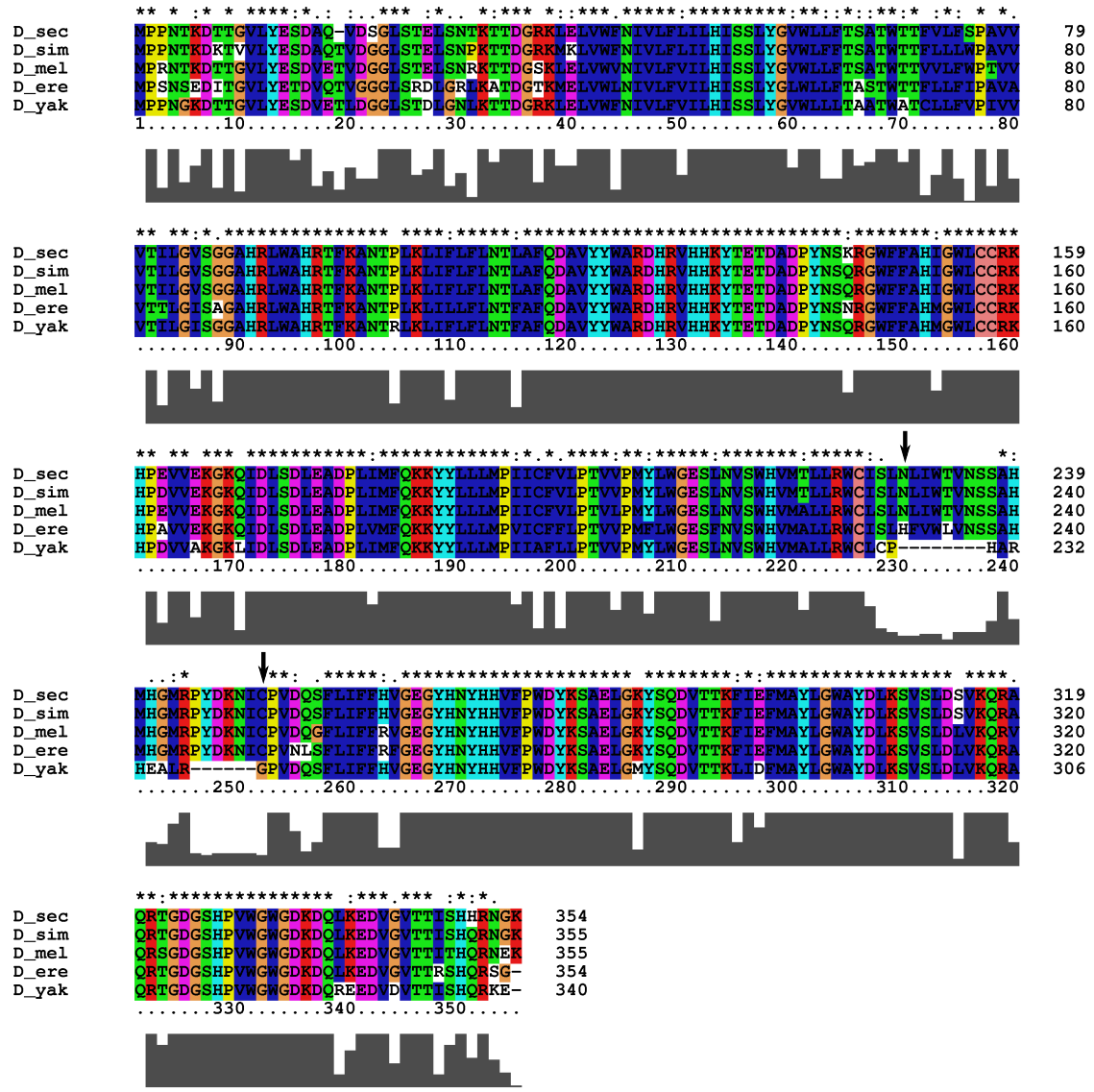


Figure 2.2: Alignment of *melanogaster* group DesatF amino acid sequences with "intron" region removed. Removing the 32 bp annotated as an intron obliterates sequence between positions 230 and 252 (indicated by black arrows), which is highly conserved in all other species.

gene tree image for desaturases (Kersey *et al.*, 2010). Sequences were aligned using MAFFT (Katoh *et al.*, 2002), using the FFT-NS-2 "Fast but rough" strategy. ML phylogenies were reconstructed using MODELGENERATOR with 4 Γ categories (substitution model chosen based on the AIC criterion) (Keane *et al.*, 2006), followed by TREEFINDER (Jobb *et al.*, 2004). Consensus trees were obtained using 250 bootstrap replicates and CONSENSE in the PHYLIP package (Felsenstein, 2005).

2.3 Results

2.3.1 Ortholog detection

Reciprocal BLAST searching discovered 96 loci among the 12 species. Six more loci were detected by MAST using the pattern from MEME, and another three loci were revealed during the final "all-against-all" BLAST search. Including the nine previously-known *D. melanogaster* loci, this brings the total number of desaturase loci from all species to 114. Details of all genes found are shown in Table 2.3. A summary of the overall results is shown in Figure 2.3. Amino acid sequences of all genes are included on the supplementary data CD.

Table 2.3: A summary of all genes found, including those of *D. melanogaster* (already known). Protein sequences are given in a separate fasta file.

Species	Gene name	Chromosome/Scaffold	No. of exons
D. ananassae v1.3	*Cyt-b5-r*	scaffold_12916	3
D. ananassae v1.3	*desat2*	scaffold_13340	4
D. ananassae v1.3	*desat2 b*	scaffold_13340	4
D. ananassae v1.3	*desatFα*	scaffold_13340	1
D. ananassae v1.3	*desatFδ*	scaffold_13340	1

Species	Gene name	Chromosome/Scaffold	No. of exons
D. ananassae v1.3	*desatFϵ*	scaffold_13337	1
D. ananassae v1.3	*desat1*	scaffold_13340	4
D. ananassae v1.3	*CG8630*	scaffold_13340	4
D. ananassae v1.3	*CG17928*	scaffold_12916	4
D. ananassae v1.3	*CG9743*	scaffold_13340	5
D. ananassae v1.3	*CG9747*	scaffold_13340	4
D. ananassae v1.3	*CG15531*	scaffold_13340	4
D. erecta v1.3	*Cyt-b5-r*	scaffold_4845	3
D. erecta v1.3	*desatFα*	scaffold_4784	1
D. erecta v1.3	*desat1*	scaffold_4770	4
D. erecta v1.3	*CG8630*	scaffold_4770	4
D. erecta v1.3	*CG17928*	scaffold_4845	4
D. erecta v1.3	*CG9743*	scaffold_4820	5
D. erecta v1.3	*CG9747*	scaffold_4820	4
D. erecta v1.3	*CG15531*	scaffold_4820	4
D. grimshawi v1.3	*Cyt-b5-r*	scaffold_15252	3
D. grimshawi v1.3	*desat2*	scaffold_15074	4
D. grimshawi v1.3	*desat1*	scaffold_15074	4
D. grimshawi v1.3	*CG8630*	scaffold_14906	4
D. grimshawi v1.3	*CG17928*	scaffold_15252	4
D. grimshawi v1.3	*CG9743*	scaffold_14906	5
D. grimshawi v1.3	*CG9747*	scaffold_14906	4
D. grimshawi v1.3	*CG15531*	scaffold_14906	4
D. mojavensis v1.3	*Cyt-b5-r*	scaffold_6500	3

Species	Gene name	Chromosome/Scaffold	No. of exons
D. mojavensis v1.3	*desat2*	scaffold_6540	4
D. mojavensis v1.3	*desat1*	scaffold_6540	4
D. mojavensis v1.3	*CG8630*	scaffold_6540	4
D. mojavensis v1.3	*CG17928*	scaffold_6500	4
D. mojavensis v1.3	*CG9743*	scaffold_6540	5
D. mojavensis v1.3	*CG9747*	scaffold_6540	4
D. mojavensis v1.3	*CG15531*	scaffold_6540	4
D. persimilis v1.3	*Cyt-b5-r*	scaffold_1	3
D. persimilis v1.3	*desat2*	scaffold_19	4
D. persimilis v1.3	*desatFα*	scaffold_47	1
D. persimilis v1.3	*desatFβ*	scaffold_0	1
D. persimilis v1.3	*desatFγ*	scaffold_3	1
D. persimilis v1.3	*desat1*	scaffold_19	4
D. persimilis v1.3	*CG8630*	scaffold_0	4
D. persimilis v1.3	*CG17928*	scaffold_1	4
D. persimilis v1.3	*CG9743*	scaffold_7	5
D. persimilis v1.3	*CG9747*	scaffold_7	4
D. persimilis v1.3	*desat1 b*	scaffold_22	3
D. persimilis v1.3	*CG15531*	scaffold_7	4
D. pseudoobscura v2.3	*Cyt-b5-r*	4_group3	3
D. pseudoobscura v2.3	*desat2*	2	4
D. pseudoobscura v2.3	*desatFα*	XR_group6	1
D. pseudoobscura v2.3	*desat1*	2	4
D. pseudoobscura v2.3	*CG8630*	2	4

Species	Gene name	Chromosome/Scaffold	No. of exons
D. pseudoobscura v2.3	*CG17928*	4_group3	4
D. pseudoobscura v2.3	*CG9743*	2	5
D. pseudoobscura v2.3	*CG9747*	2	4
D. pseudoobscura v2.3	*CG15531*	2	4
D. pseudoobscura v2.3	*desat1 b*	XL_group1e	3
D. pseudoobscura v2.3	*desatFβ*	2	1
D. pseudoobscura v2.3	*desatFγ*	2	1
D. sechellia v1.3	*Cyt-b5-r*	scaffold_7	3
D. sechellia v1.3	*desat2*	scaffold_0	4
D. sechellia v1.3	*desatFα*	scaffold_0	1
D. sechellia v1.3	*desat1*	scaffold_0	4
D. sechellia v1.3	*CG8630*	scaffold_0	4
D. sechellia v1.3	*CG17928*	scaffold_7	4
D. sechellia v1.3	*CG9743*	scaffold_4	5
D. sechellia v1.3	*CG9747*	scaffold_4	4
D. sechellia v1.3	*CG15531*	scaffold_4	4
D. simulans v1.3	*Cyt-b5-r*	2L	3
D. simulans v1.3	*desat2*	3R	4
D. simulans v1.3	*desatFα*	3L	1
D. simulans v1.3	*CG8630*	3R	4
D. simulans v1.3	*CG17928*	2L	4
D. simulans v1.3	*CG9743*	3R	5
D. simulans v1.3	*CG9747*	3R	4
D. simulans v1.3	*desat1*	3R	4

Species	Gene name	Chromosome/Scaffold	No. of exons
D. simulans v1.3	*CG15531*	3R	4
D. virilis v1.2	*Cyt-b5-r*	scaffold_12963	3
D. virilis v1.2	*desat2*	scaffold_13047	4
D. virilis v1.2	*desat1*	scaffold_13047	4
D. virilis v1.2	*CG8630*	scaffold_12855	4
D. virilis v1.2	*CG17928*	scaffold_12963	4
D. virilis v1.2	*CG9743*	scaffold_12855	5
D. virilis v1.2	*CG9747*	scaffold_12855	4
D. virilis v1.2	*CG15531*	scaffold_12855	4
D. willistoni v1.3	*Cyt-b5-r*	scf2_1100000004577	3
D. willistoni v1.3	*desat2*	scf2_1100000004943	3
D. willistoni v1.3	*desatFζ*	scf2_1100000004511	1
D. willistoni v1.3	*desatFη*	scf2_1100000004762	1
D. willistoni v1.3	*desat1*	scf2_1100000004943	4
D. willistoni v1.3	*CG8630*	scf2_1100000004943	4
D. willistoni v1.3	*CG17928*	scf2_1100000004577	4
D. willistoni v1.3	*CG9743*	scf2_1100000004943	5
D. willistoni v1.3	*CG9747*	scf2_1100000004943	4
D. willistoni v1.3	*CG15531*	scf2_1100000004943	4
D. yakuba v1.3	*Cyt-b5-r*	2R	3
D. yakuba v1.3	*desat2*	3R	4
D. yakuba v1.3	*desatFα*	3L	1
D. yakuba v1.3	*desat1*	3R	4
D. yakuba v1.3	*CG8630*	3R	4

Species	Gene name	Chromosome/Scaffold	No. of exons
D. yakuba v1.3	*CG17928*	2R	4
D. yakuba v1.3	*CG9743*	3R	5
D. yakuba v1.3	*CG9747*	3R	4
D. yakuba v1.3	*CG15531*	3R	4
D. melanogaster v5.13	*CG17928*	2L	4
D. melanogaster v5.13	*CG8630*	3R	4
D. melanogaster v5.13	*CG9743*	3R	5
D. melanogaster v5.13	*CG9747*	3R	4
D. melanogaster v5.13	*Cyt-b5-r*	2L	3
D. melanogaster v5.13	*desat1*	3R	4
D. melanogaster v5.13	*desat2*	3R	4
D. melanogaster v5.13	*desatFα*	3L	1
D. melanogaster v5.13	*CG15531*	3R	4

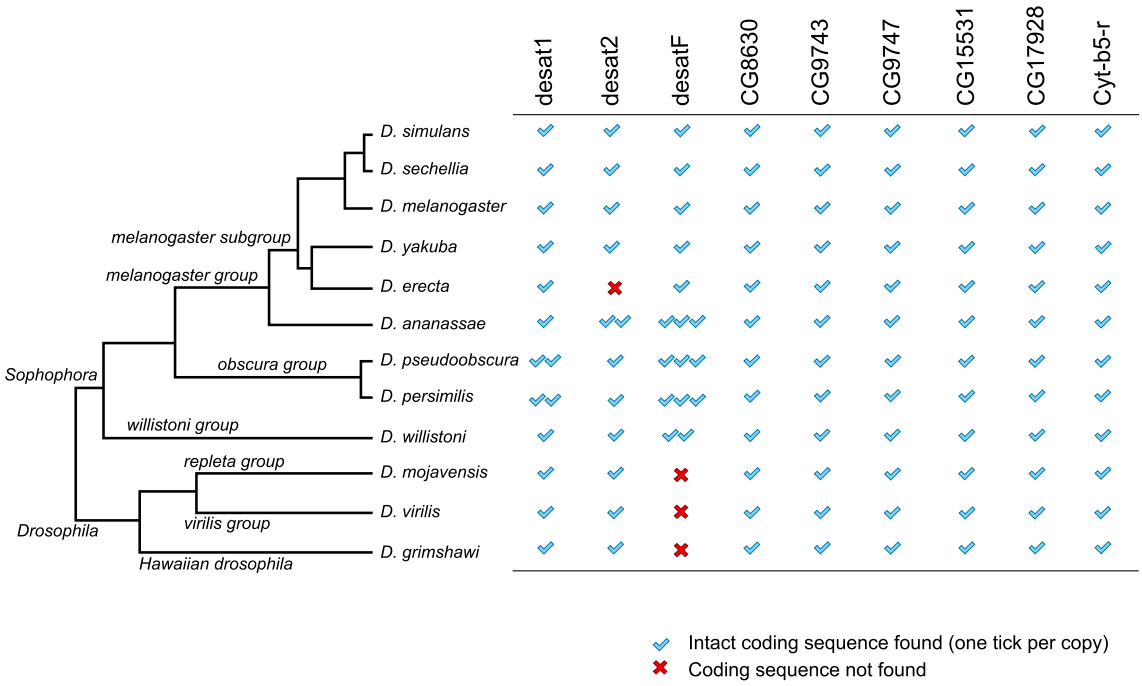


Figure 2.3: Summary of ortholog search results

DesatF was only found in species of the *Sophophora* subgenus, and not in *D. mojavensis*, *D. virilis* or *D. grimshawi*. This gene has been duplicated the most:

twice in *D. ananassae* and the *obscura* group species, and once in *D. willistoni*. The desatF orthologs found in our study correspond to those recently described by Fang *et al.* (2009) using homology-based searching and synteny analysis, and have been labelled in the same way. The only other gene loss here is *desat2* from *D. erecta*, which was also noted by Fang *et al.* (2009).

Additional duplication events in *desat1* and *desat2* were also identified. *desat2* has been duplicated once in *D. ananassae*. The two species of the *obscura* group, *D. persimilis* and *D. pseudoobscura*, have a second copy of *desat1*, which has been named *desat1b*. This has fewer exons (three instead of four) than the other ancestral *desat1* gene. Exon 1 appears to be a fusion of exons 1 and 2 of *desat1*, while exons 2 and 3 correspond to exons 3 and 4 of *desat1*. The first 56 amino acids of the protein sequence of exon 1 in *D. pseudoobscura desat1b* have only 31.4% identity with the same region of *desat1* in the same species (Figure 2.4). The protein sequences of Desat1 and Desat1b were analysed using SIGNALP (Bendtsen et al. 2004) to check for the presence of signal peptides, but neither sequence showed evidence for these. Amino acid identities between these, and the other sets of duplicates are modest, indicating up to 37% divergence (Table 2.4). All other genes are present in a single copy in all species.

Species	Gene	Percent Identity
D. ananassae	*Desat2*	67.6
	DesatF	71.5
D. pseudoobscura	*Desat1*	76.5
	DesatF	71.9
D. persimilis	*Desat1*	76.3
	DesatF	72.8
D. willistoni	*DesatF*	62.7

Table 2.4: The percentage identities of multiple sequence alignments between the amino acid sequences of each set of duplicates.

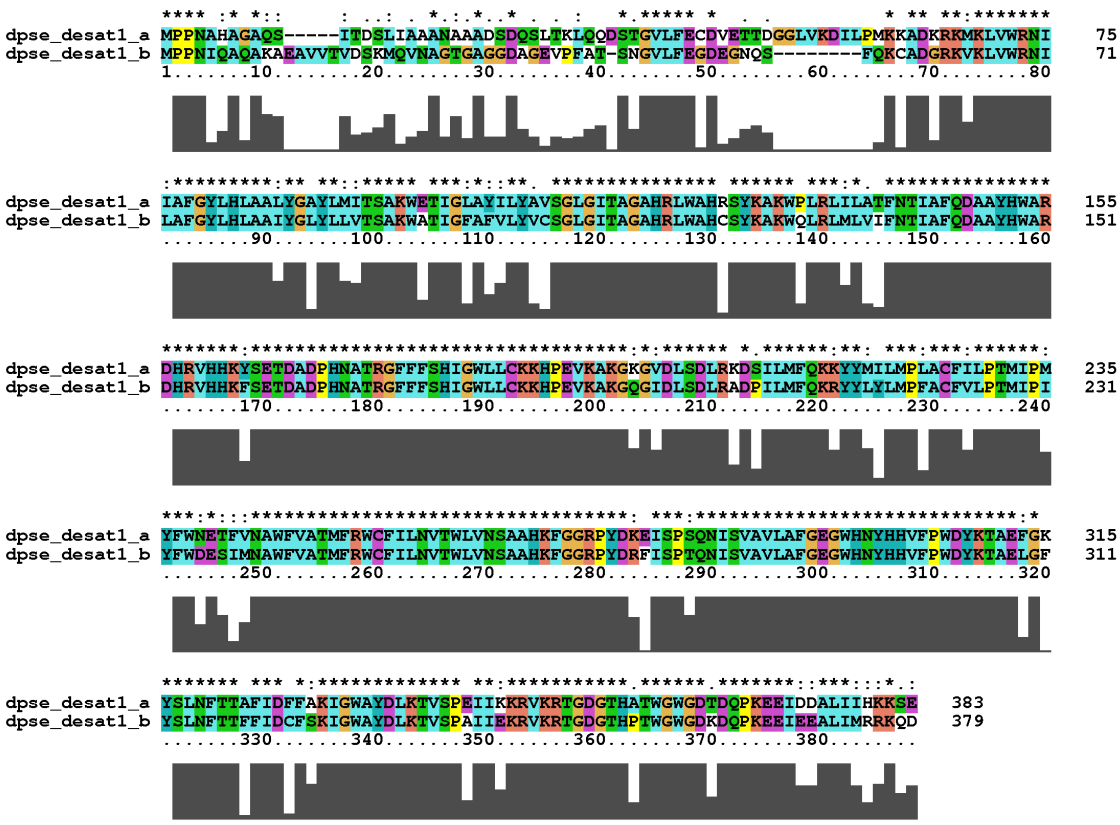


Figure 2.4: Alignment of *D. pseudoobscura* Desat1 and Desat1b amino acid sequences. The alignment shows that the N-terminal region of Desat1b is highly divergent, compared with the remainder of the sequence. Alignment created using MAFFT (Katoh *et al.*, 2002), figure created using ClustalX (Thompson *et al.*, 1994).

2.3.2 Re-sequencing

The resequencing of *D. yakuba desatF* in strain 14021-0261.00 showed that this strain does not have a deletion, and the coding sequence is therefore intact. The strain used in the genome project, 14021-0261.01, does have the deletion, however. *D. sechellia desat1* was found to have an intact coding sequence.

2.3.3 Phylogenetics

Nucleotide and amino acid phylogenies were derived for all loci across all species, to further examine the evolution of the gene family (Figures 2.5 to 2.8). The amino acid consensus phylogeny (Figure 2.5) suggests that an initial duplication event produced the ancestor of *desat1* and *desat2*, and the ancestor of the rest of the desaturases. Next, the *CG17928-Cyt-b5-r* clade arose, and after this a retrotransposition event occurred which gave rise to *desatF*. It was after this that the gene from which the retrocopy was duplicated then duplicated again three more times, to produce *CG8630*, *CG9743*, *CG9747* and *CG15531*. This interpretation is illustrated in Figure 2.9.

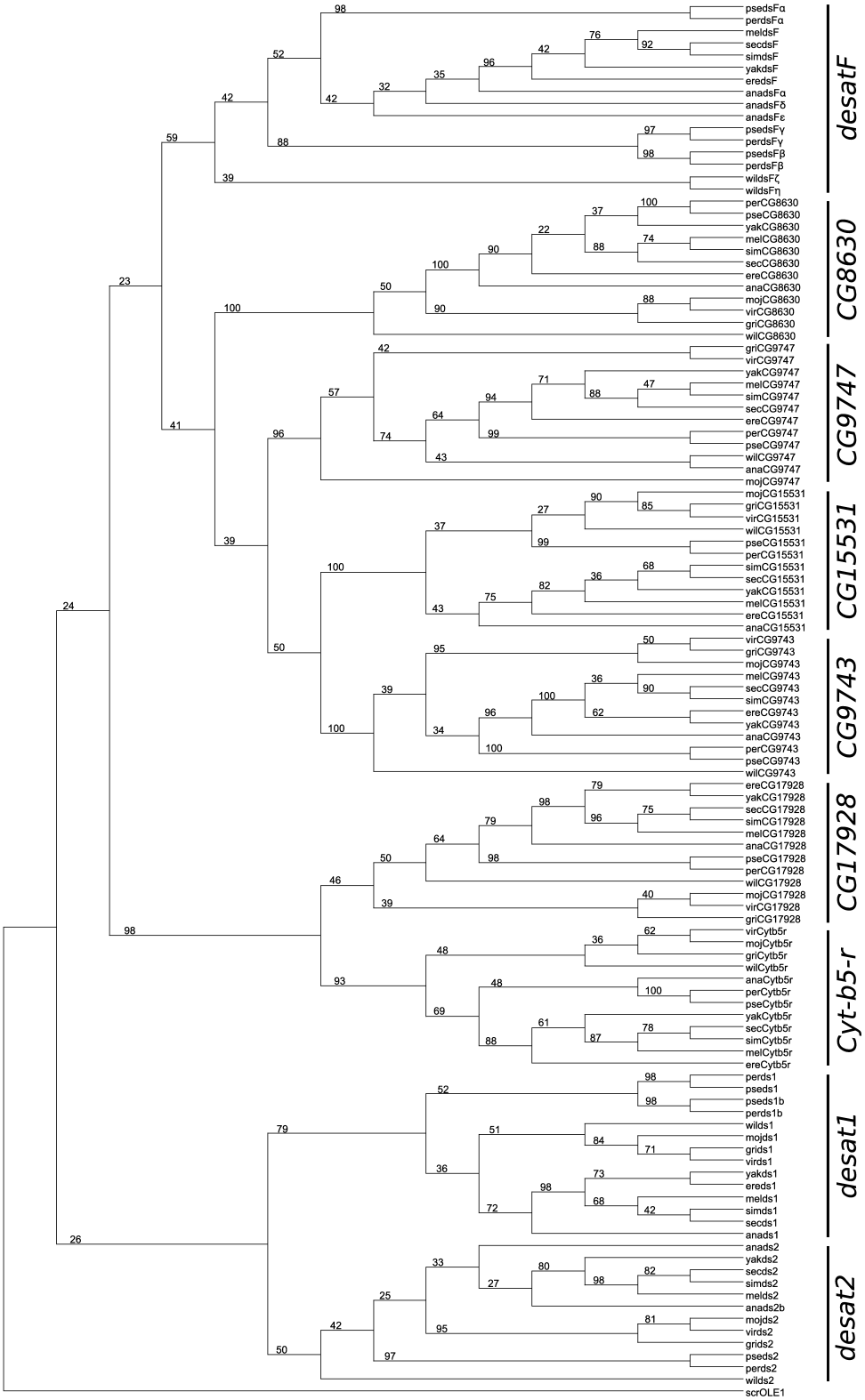


Figure 2.5: Consensus phylogeny of all desaturase amino acid sequences found and *S. cerevisiae* outgroup (scrOLE1), reconstructed using LG+I+G model. Percentage bootstrap support is indicated. Gene names are given as abbreviated species name followed by *D. melanogaster* gene name. Desat1, Desat2 and DesatF abbreviated to ds1, ds2 and dsF respectively.

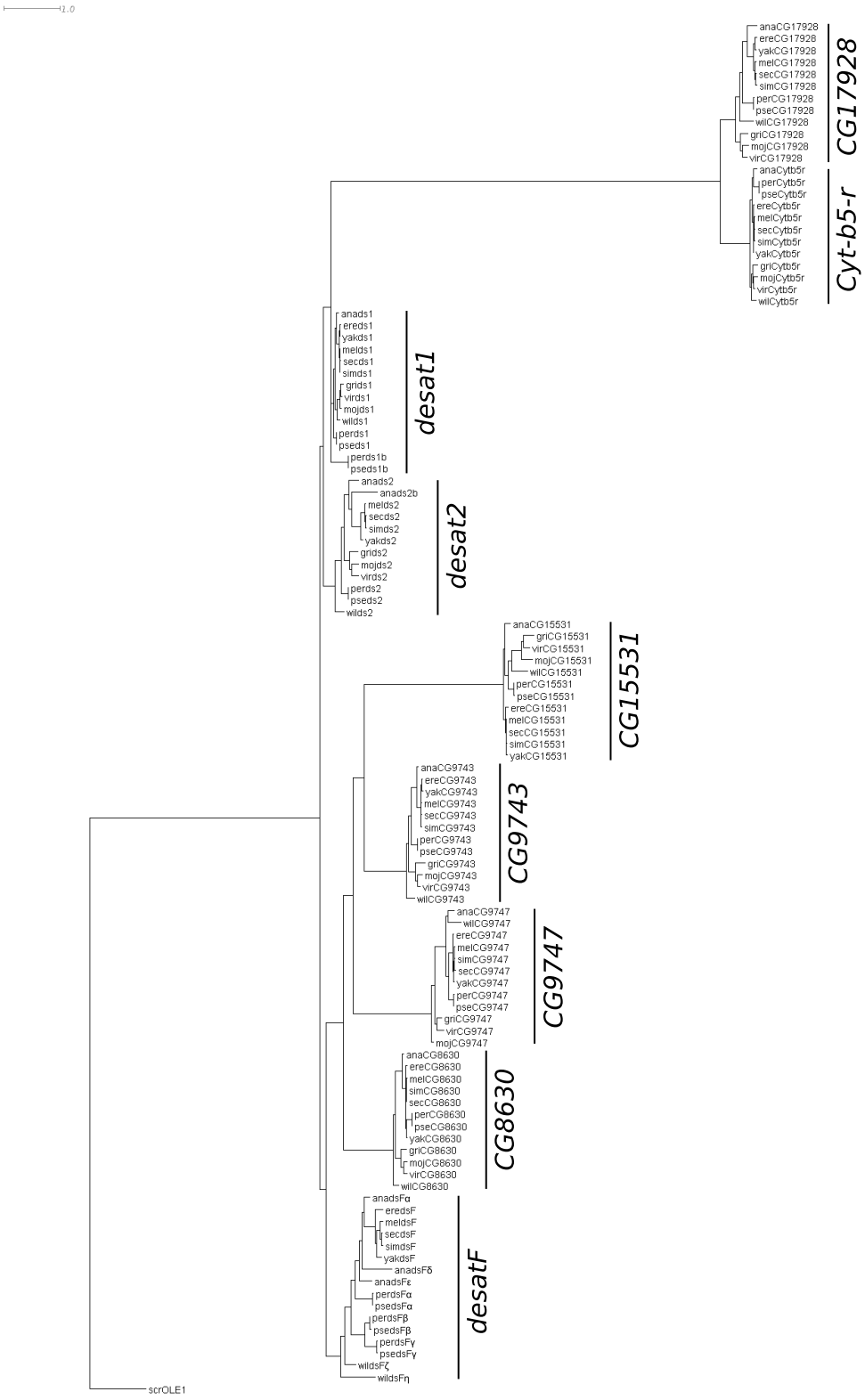


Figure 2.6: Maximum likelihood phylogeny of amino acid sequences showing branch lengths (LG+I+G model). Gene names are given as abbreviated species name followed by *D. melanogaster* gene name. Desat1, Desat2 and DesatF abbreviated to ds1, ds2 and dsF respectively. Scale bar indicates expected substitutions per residue.

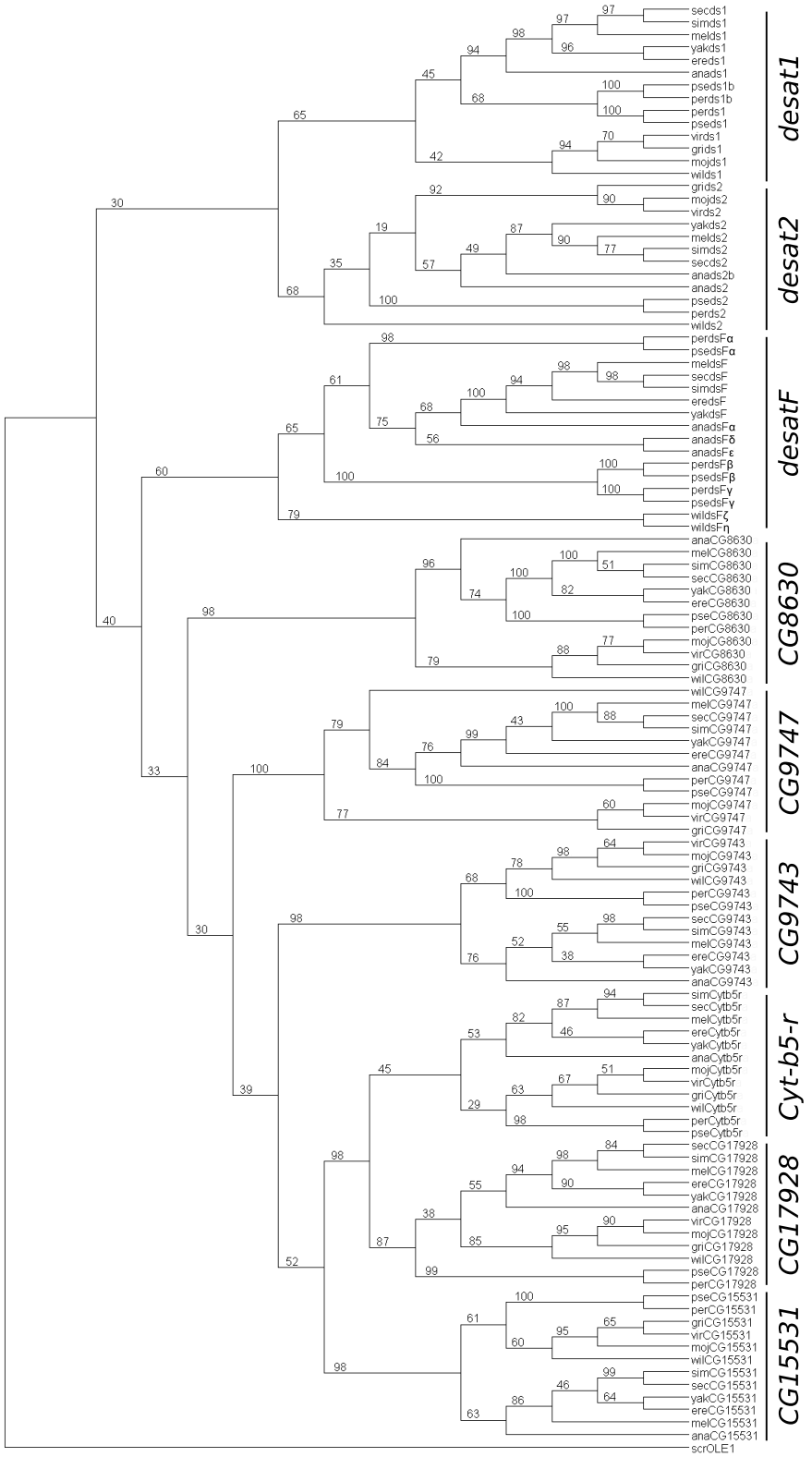


Figure 2.7: Consensus phylogeny of protein-coding nucleotide sequences of all desaturases found and *S. cerevisiae* outgroup (scrOLE1), reconstructed using GTR+I+G model. Percentage bootstrap support is indicated. Gene names are given as abbreviated species name followed by *D. melanogaster* gene name. Desat1, Desat2 and DesatF abbreviated to ds1, ds2 and dsF respectively.

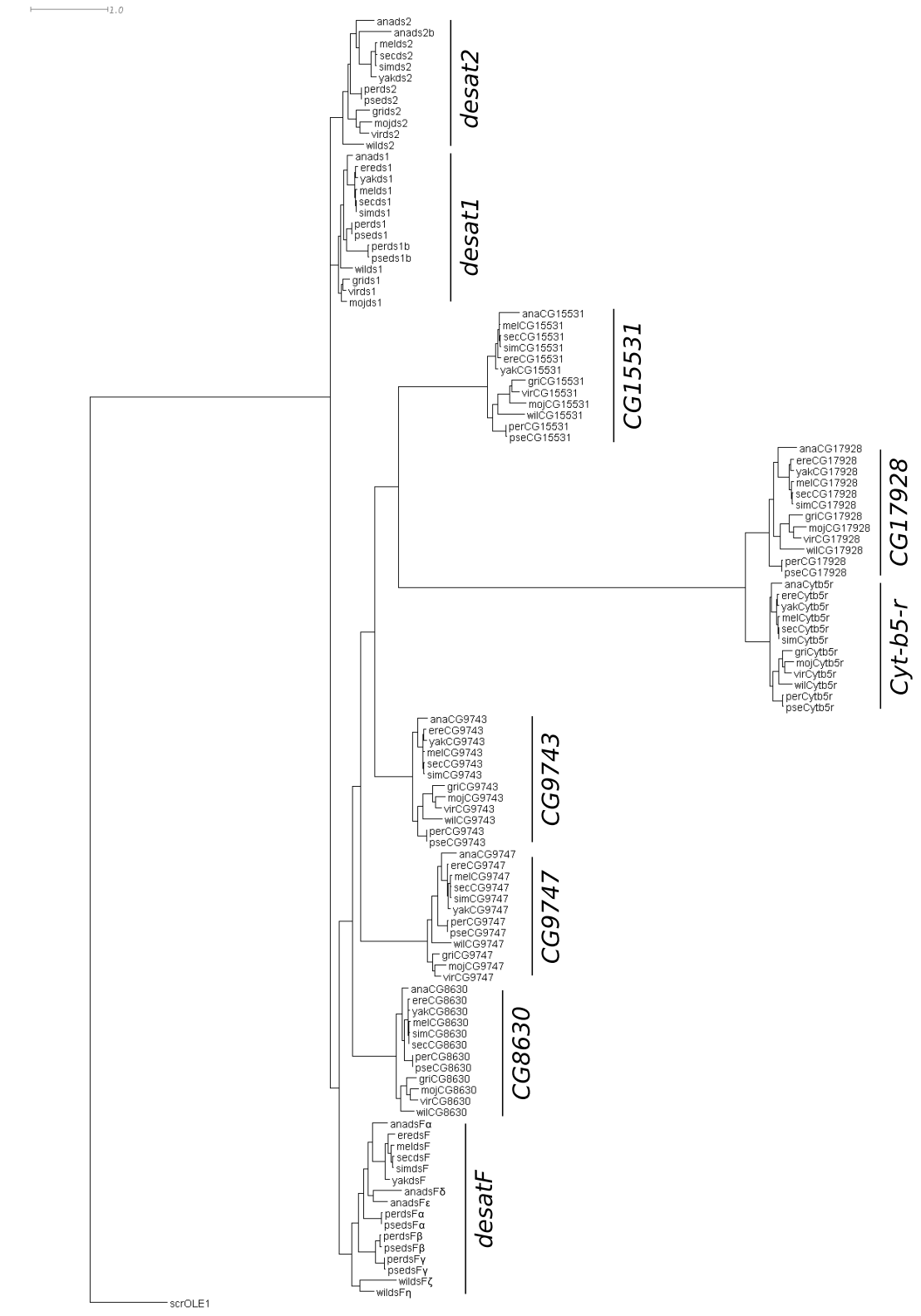


Figure 2.8: Maximum likelihood phylogeny of protein-coding nucleotide sequences showing branch lengths (GTR+I+G model). Gene names are given as abbreviated species name followed by *D. melanogaster* gene name. Desat1, Desat2 and DesatF abbreviated to ds1, ds2 and dsF respectively. Scale bar indicates expected substitutions per residue.

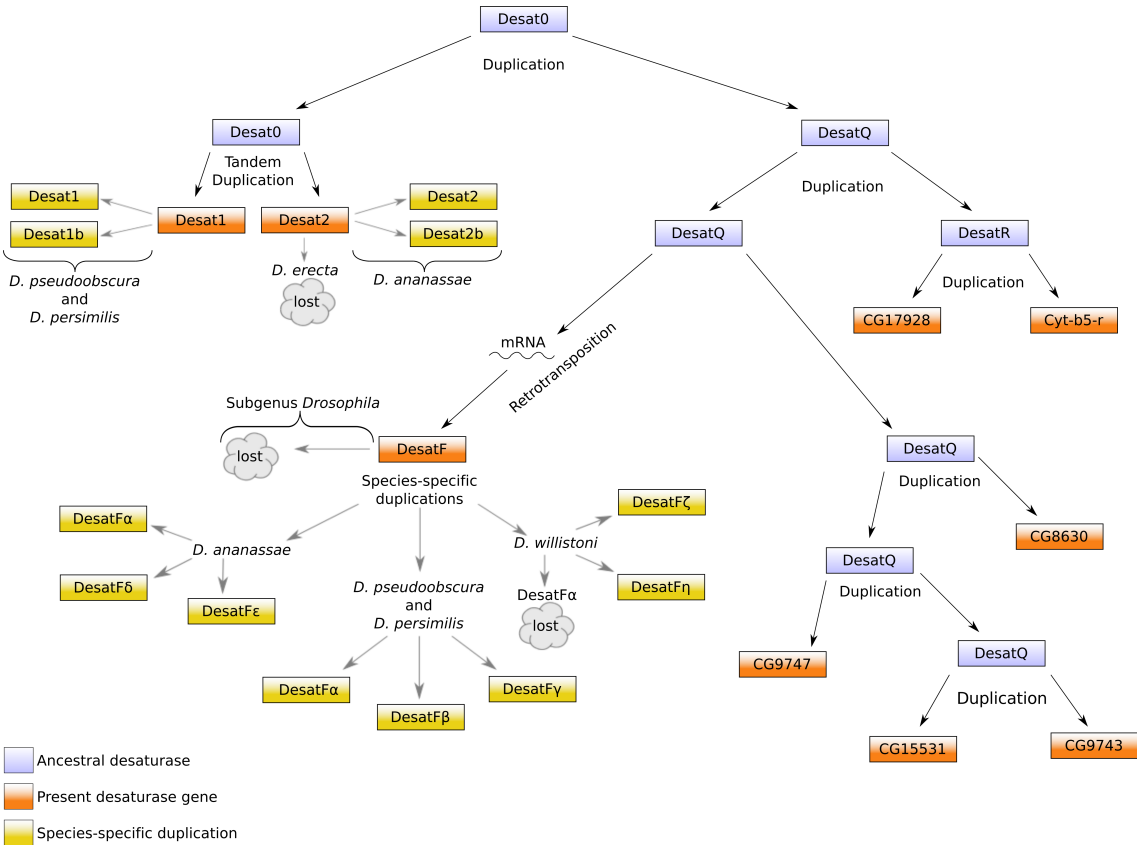


Figure 2.9: Hypothesis for the evolution of the desaturase genes in *Drosophila*. Proposed ancestral genes (pale blue) were arbitrarily named *desat0*, *desatQ* and *desatR*. *DesatF* was duplicated many times in various *Sophophora* species. The original *DesatF* gene is labelled *DesatFα* in the species-specific duplicates. This diagram is based on the consensus phylogeny of desaturase amino acid sequences (Figure 2.5), which contains some very low bootstrap support values. Many of the key nodes cannot be confidently resolved, and hence this diagram remains a hypothesis of the evolution of the desaturase gene family, rather than a concrete conclusion.

2.4 Discussion

This study shows that the majority of desaturase loci are present in a single copy in each of the 12 sequenced species of *Drosophila*. The striking exceptions to this, however, are the three loci known to be involved in pheromone biosynthesis: these loci show a large amount of duplication and loss across the species. The most duplicated gene is *desatF*, which is present in multiple copies in four species (Figure 2.3). This, coupled with the observations of Shirangi *et al.* (2009) which demonstrate rapid evolution of sex-specific expression variation in many species, make this gene an interesting candidate for further investigation.

DesatF has also been lost from the subgenus *Drosophila*. On first look at the results in Figure 2.3, it is tempting to conclude that the *desatF* gene must have arisen after or during the split between the *Sophophora* and *Drosophila*. However, taking into account the phylogenetic analyses performed here, this cannot have been the case. They show that *desatF* arose before *CG8630*, *CG9747*, *CG9743* and *CG15531*. These four genes are present in all species studied. Therefore, they must have arisen in a single common ancestral species, before any of the species seen today existed. After this, the ancestors of the *Sophophora* and *Drosophila* diverged, and presumably the *Drosophila* ancestor lost its copy of *desatF*. This conclusion concurs with that of Fang *et al.* (2009), and the search methods utilised here discovered all loci found in their study. As well as finding all desaturase loci previously identified, the methods used here also uncovered two more loci: a putative duplicate of *desat2* in *D. ananassae*, and one of *desat1* in *D. pseudoobscura* and *D. persimilis*. Nothing is yet known about the function of these genes, but they have intact coding sequences and are closely related to genes with previously demonstrated functions in pheromone modification. In the phylogenetic analysis by Fang *et al.* (2009), *CG8630* is used as an outgroup to *desat1*, *desat2* and *desatF*. This difference does not affect the conclusions regarding the evolution of these three genes, however. The analyses in this chapter show that *CG8630* is not a natural outgroup to these genes, as it appears to have arisen later than *desat1*, *desat2* and *desatF* (Figures 2.5 and 2.9). It should be noted that the bootstrap support values of the consensus phylogenies shown in this chapter (Figures 2.5 and 2.7) are relatively low, and it is therefore possible that

the real topology of the desaturase phylogeny could differ from this.

Similar studies of *Drosophila* gene family evolution have been completed since the release of the 12 genomes in 2007. Hahn *et al.* (2007) conducted an analysis of gain and loss of loci across the whole genome of all 12 species. They found that 41% of all 11434 gene families have changed in size in at least one species, and some cases of extreme gene family expansion or contraction, particularly in genes related to sex and spermatogenesis. Vieira *et al.* (2007) studied the odorant binding protein (OBP) family, which contains proteins that make contact with odorant molecules and are thought to initiate chemosensory pathways. They found that the OBP family contains between 40 and 61 genes across the 12 species, with in total 43 gains and 28 losses. Guo and Kim (2007) performed a similar analysis of olfactory receptor genes, a large family in which they report 59 orthologous groups and a substantial amount of gain and loss, especially in *D. willistoni*, *D. grimshawi* and the *obscura* group. The desaturase gene family is smaller than both of these, and does not appear to have undergone extreme expansion or contraction in any of the 12 species; although the *desatF* gene has been duplicated independently several times, it is at most present in three copies, in *D. ananassae* and the *obscura* group.

The work presented in this chapter has uncovered many desaturase loci with intact coding sequences, across the 12 species. It is not possible to tell from this work, however, whether any of the novel genes discovered have a role in pheromone modification. Although all genes located have intact coding sequences, it is not possible to discern whether they are actually functional. A step towards discerning those, if any, that may have interesting evolutionary stories, is to look for signatures of natural selection within their sequences. This is investigated in the next chapter.

Bibliography

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**:403–410.

Bailey, T. L., Williams, N., Misleh, C. and Li, W. W. (2006). MEME: discovering and analysing DNA and protein sequence motifs. *Nucleic Acids Research*, **34**:W369–W373.

Billeter, J. C., Atallah, J., Krupp, J. J. and Levine, J. D. (2009). Specialized cells tag sexual and species identity in *Drosophila melanogaster*. *Nature*, **461**:987–991.

Birney, E., Clamp, M. and Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, **14**:988–995.

Blows, M. W. and Alan, R. A. (1998). Levels of mate recognition within and between two *Drosophila* species and their hybrids. *The American Naturalist*, **152**:826–837.

Bousquet, F., Houot, B., Chauvel, I., Dupas, S. and Ferveur, J. F. (2009). *desat1* and the evolution of pheromonal communication in *Drosophila*. *Annals of the New York Academy of Sciences*, **1170**:502–505.

Butler, B., Pirrotta, V., Irminger-Finger, I. and Nöthiger, R. (1987). The sex-determining *tra* gene of *Drosophila*: molecular cloning and transformation studies. *The EMBO Journal*, **5**:3607–3613.

Chertemps, T., Duportets, L., Labeur, C., Ueyama, M. and Wicker-Thomas, C. (2006). A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in *Drosophila melanogaster*. *Insect Molecular Biology*, **15**:465–473.

Cobb, M. and Jallon, J.-M. (1990). Pheromones, mate recognition and courtship stimulation in the *Drosophila melanogaster* species sub-group. *Animal Behaviour*, **39**:1058–1067.

Coyne, J. A. (1996). Genetics of differences in pheromonal hydrocarbons between *Drosophila melanogaster* and *D. simulans*. *Genetics*, **143**:353–364.

Coyne, J. A., Crittenden, A. P. and Mah, K. (1994). Genetics of a pheromonal difference contributing to reproductive isolation in *Drosophila*. *Science*, **265**:1461–1464.

Coyne, J. A. and Elwyn, S. (2006a). Does the *desaturase2* locus in *Drosophila melanogaster* cause adaptation and sexual isolation? *Evolution*, **60**:279–291.

Coyne, J. A. and Elwyn, S. (2006b). *Desaturase-2*, environmental adaptation, and sexual isolation in *Drosophila melanogaster*. *Evolution*, **60**:626–627.

Coyne, J. A., Wicker-Thomas, C. and Jallon, J.-M. (1999). A gene responsible for a cuticular hydrocarbon polymorphism in *Drosophila melanogaster*. *Genetical Research*, **73**:189–203.

Dallerac, R., Labeur, C., Jallon, J.-M., Knipple, D. C., Roelofs, W. L. and Wicker-Thomas, C. (2000). A Δ9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **97**:9449–9454.

Doi, M., Nemoto, T., Nakanishi, H., Kurahara, Y. and Oguma, Y. (1997). Behavioral response of males to major sex pheromone component, (z,z)-5,25-hentriacontadiene, of *Drosophila ananassae* females. *Journal of Chemical Ecology*, **23**:2067–2077.

Droney, D. and Hock, B. (1998). Male sexual signals and female choice in *Drosophila grimshawi* (Diptera: Drosophilidae). *Journal of Insect Behavior*, **11**:59–71.

Duffy, J. B. (2002). GAL4 system in *Drosophila*: a fly geneticist's swiss army knife. *Genesis*, **34**:1–15.

Erdman, S. E., Chen, H. J. and Burtis, K. C. (1996). Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila* doublesex proteins. *Genetics*, **144**:1639–1652.

Etges, W. J. and Ahrens, M. A. (2001). Premating isolation is determined by larval-rearing substrates in cactophilic *Drosophila mojvensis*. v. deep geographic variation in epicuticular hydrocarbons among isolated populations. *The American Naturalist*, **158**:585–598.

Etges, W. J. and Jackson, L. (2001). Epicuticular hydrocarbon variation in *Drosophila mojavensis* cluster species. *Journal of Chemical Ecology*, **27**:2125–2149.

Fang, S., Ting, C.-T., Lee, C.-R., Chu, K.-H., Wang, C.-C. and Tsaur, S.-C. (2009). Molecular evolution and functional diversification of fatty acid desaturases after recurrent gene duplication in *Drosophila*. *Molecular Biology and Evolution*, **26**:1447–1456.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package).

Ferveur, J. F. (2005). Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behavior Genetics*, **35**:269–295.

Ferveur, J.-F., Cobb, M., Boukella, H. and Jallon, J.-M. (1996). World-wide variation in *Drosophila melanogaster* sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. *Genetica*, **97**:73–80.

Ferveur, J. F., Savarit, F., O'Kane, C. J., Sureau, G., Greenspan, R. J. and Jallon, J. M. (1997). Genetic feminization of pheromones and its behavioral consequences in *Drosophila* males. *Science*, **276**:1555–1558.

Gleason, J. M., Jallon, J.-M., Rouault, J.-D. and Ritchie, M. G. (2005). Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between *Drosophila simulans* and *D. sechellia*. *Genetics*, **171**:1789–1798.

Greenberg, A. J., Moran, J. R., Coyne, J. A. and Wu, C.-I. (2003). Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science*, **302**:1754–1757.

Greenspan, R. J. and Ferveur, J. F. (2000). Courtship in *Drosophila*. *Annual Review of Genetics*, **34**:205–32.

Grillet, M., Dartevelle, L. and Ferveur, J. F. (2006). A *Drosophila* males pheromone affects female sexual receptivity. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **273**:315–323.

Guo, S. and Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular Biology and Evolution*, **24**:1198–1207.

Hahn, M. W., Han, M. V. and Han, S. G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, **3**:e197.

Hall, J. C. (1994). The mating of a fly. *Science*, **264**:1702–1714.

Houot, B., Bousquet, F. and Ferveur, J. F. (2010). The consequences of regulation of desat1 expression for pheromone emission and detection in *Drosophila melanogaster*. *Genetics*, **185**:1297–1309.

Jallon, J.-M. (1984). A few chemical words exchanged by *Drosophila* during courtship and mating. *Behavior Genetics*, **14**:441–478.

Jallon, J. M. and David, J. R. (1987). Variations inf cuticular hydrocarbons among the eight species of the *Drosophila melanogaster* subgroup. *Evolution*, **41**:294–302.

Jobb, G., von Haeseler, A. and Strimmer, K. (2004). TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology*, **4**:18.

Karlson, P. and Butenandt, A. (1959). Pheromones (Ectohormones) In Insects. *Annual Review of Entomology*, **4**:39–58.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**:3059–3066.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. M. and McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology*, **6**:29.

Keays, M. C., Barker, D., Wicker-Thomas, C. and Ritchie, M. G. (2011). Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. *Molecular Ecology*, **20**:3617–3630.

Kersey, P. J., Lawson, D. *et al.* (2010). Ensembl Genomics: extending Ensembl across the taxonomic space. *Nucleic Acids Research*, **38**:D563–D569.

Legendre, A., Miao, X. X., Lage, J. L. D. and Wicker-Thomas, C. (2008). Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in *Drosophila*. *Insect Biochemistry and Molecular Biology*, **38**:244–255.

Liimatainen, J. O. and Hoikkala, A. (1998). Interactions of the males and females of three sympatric *Drosophila virilis*-group species, *D. montana*, *D. littoralis* and *D. lummei*, (Diptera: Drosophilidae) in intra- and interspecific courtships in the wild and in the laboratory. *Journal of Insect Behvior*, **11**:399–417.

Liimatainen, J. O. and Jallon, J. M. (2007). Genetic analysis of cuticular hydrocarbons and their effect on courtship in *Drosophila virilis* and *D. lummei*. *Behavior Genetics*, **37**:713–725.

Marcillac, F., Bousquet, F., Alabouvette, J., Savarit, F. and Ferveur, J. F. (2005a). A mutation with major effects on *Drosophila melanogaster* sex pheromones. *Genetics*, **171**:1617–1628.

Marcillac, F., Grosjean, Y. and Ferveur, J. F. (2005b). A single mutation alters production and discrimination of *Drosophila* sex pheromones. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **272**:303–309.

Michalak, P., Malone, J. H., Lee, I. T., Hoshino, D. and Ma, D. (2007). Gene expression polymorphism in *Drosophila* populations. *Molecular Ecology*, **16**:1179–1189.

Nakamura, M. T. and Nara, T. Y. (2004). Structure, function and dietary regulation of $\Delta 6$, $\Delta 5$, and $\Delta 9$ desaturases. *Annual Review of Nutrition*, **24**:345–376.

Roelofs, W. and Bjostad, L. (1984). Biosynthesis of Lepidopteran Pheromones. *Bioorganic Chemistry*, **12**:279–298.

Roelofs, W. L. and Rooney, A. P. (2003). Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *PNAS*, **100**:9179–9184.

Rouault, J., Capy, P. and Jallon, J.-M. (2000). Variations of male cuticular hydrocarbons with geoclimatic variables: an adaptive mechanism in *Drosophila melanogaster*? *Genetica*, **110**:117–130.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S. and Misener, S. (eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, pages 365–386. Totowa, New Jersey: Humana Press.

Savarit, F., Sureau, G., Cobb, M. and Ferveur, J. F. (1999). Genetic elimination of known pheromones reveals the fundamental chemical bases of mating and isolation in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA*, **96**:9015–9020.

Shirangi, T. R., Dufour, H. D., Williams, T. M. and Carroll, S. B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biology*, **7**:e1000168.

Takahashi, A. and Ting, C.-T. (2004). Genetic basis of sexual isolation in *Drosophila melanogaster*. *Genetica*, **120**:273–284.

Takahashi, A., Tsaur, S.-C., Coyne, J. A. and Wu, C.-I. (2001). The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **98**:3920–3925.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, **450**:203–218.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**:4673–4680.

Ting, C.-T., Takahashi, A. and Wu, C.-I. (2001). Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. *Proceedings of the National Academy of Sciences of the USA*, **98**:6709–6713.

Vieira, F. G., Sánchez-Garcia, A. and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology*, **8**:R235.

Wicker-Thomas, C., Henriet, C. and Dallerac, R. (1997). Partial characterization of a fatty acid desaturase gene in *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology*, **27**:963–972.

Wilson, E. O. and Bossert, W. H. (1963). Chemical Communication among Animals. *Recent Progress in Hormone Research*, **19**:673–716.

Wilson, R. J., Goodman, J. L., Strelets, V. B. and the FlyBase Consortium (2008). FlyBase: integration and improvements to query tools. *Nucleic Acids Research*, **36**:D588–D593.

Wu, C.-I., Hollocher, H., Begun, D. J., Aquadro, C. F., Xu, Y. and Wu, M.-L. (1995). Sexual isolation in *Drosophila melanogaster*: A possible case of incipient speciation. *Proceedings of the National Academy of Sciences of the USA*, **92**:2519–2523.

Chapter 3

Signatures of Selection in *Drosophila* Desaturase Loci

3.1 Introduction

3.1.1 Positive selection and species evolution

When changes to an amino acid sequence are advantageous, they have a higher probability of surviving and becoming fixed in the population. Positive selection promotes the evolution of advantageous mutations, and can lead to neofunctionalisation. It is essentially the opposite of negative, or purifying selection, in which any change to an amino acid sequence has deleterious or disadvantageous consequences for the organism. As such, they are not tolerated, and the amino acid sequence remains the same. Loci that are thought to play a role in adaptation and reproductive isolation between species, and therefore to be candidate speciation genes, have been found to show rapid evolution, and to evolve as a result of positive selection (Nosil and Schluter, 2011; Orr, 2005; Wu and Ting, 2004). It is therefore of interest to investigate potential speciation genes with a view to discovering whether they have been subject to positive selection.

3.1.2 How to detect positive selection in coding sequences

Due to the degeneracy of the genetic code, mutations in protein-coding nucleotide sequences may be "synonymous", meaning there is no change to the resulting amino acid sequence. A mutation in the nucleotide sequence which in turn leads to a change in the encoded protein is known as a "nonsynonymous" mutation. Selection pressure can be measured by examining the relationship between the synonymous and nonsynonymous substitution rates, d_N and d_S, using the ratio $\omega = d_N/d_S$. d_N is the number of nonsynonymous subsititutions per nonsynonymous site, and d_S is the number of synonymous substitutions per synonymous site. Under neutrality, where natural selection has no effect on fitness, the rates of synonymous and nonsynonymous substitutions are expected to be equal: $d_N = d_S$, $\omega = 1$. Under purifying selection, where mutations in the amino acid sequence are deleterious, the rate of synonymous substitutions is expected to be higher than that of nonsynonymous substitutions: $d_N < d_S$, $\omega < 1$. In cases where changes to the amino acid sequence confer an adaptive advantage and are thus favoured by natural selection, the expectation is that the nonsynonymous mutation rate will be higher than the synonymous one: $d_N > d_S$, $\omega > 1$. If ω is significantly greater than 1, this is evidence for positive selection (Yang, 2006). This method is not without problems however; although synonymous mutations are generally accepted to be neutrally evolving because they do not affect the amino acid sequence, this is not always true. In highly expressed genes, there is often strong preference for certain codons, and this has been shown to cause lower estimates of d_S (Sharp and Li, 1989; Tamura *et al.*, 2004).

Over the last few decades, mathematical methods have been developed to estimate d_N and d_S among related sequences. Miyata and Yasunaga (1980) and Perler *et al.* (1980) presented methods for measuring the rates of synonymous and nonsynonymous substitutions from alignments of homologous nucleotide sequences. Li *et al.* (1985) further developed these methods, addressing the difference between transition and transversion rates (Kimura, 1980), and multiple hits – where the actual number of changes at a particular site is underestimated. Nei and Gojobori (1986) put forward another similar method, which was simpler than those preceding it but produced essentially the same estimates of d_N and d_S.

Earlier studies implementing these methods tended to average the estimates of d_N and d_S over all codon sites in the alignment, as well as over the divergence time between the sequences. This approach is very conservative when looking for positive selection, given that most sites in a functional protein are likely to be under strong negative selection most of the time. Positive selection is thought to act over relatively short evolutionary time periods, and to only affect a small number of sites in the sequence (Yang, 2006). Therefore, these methods may fail to detect such instances of positive selection.

3.1.3 Statistical tests of positive selection

Three types of statistical test are commonly used to detect positive selection in multiple alignments of homologous sequences. Maximum likelihood models of evolution, with and without positive selection, are fitted to the sequence and phylogeny data provided. Their log-likelihood scores (ℓ) are then compared in a likelihood ratio test (LRT), to determine whether a model including positive selection is a statistically significantly better fit to the data than one that does not (Yang, 2006).

Branch tests of positive selection are used to look for selection affecting individual branches on a phylogeny (Yang, 1998; Yang and Nielsen, 1998). The basic branch model is the "one-ratio" model, in which all branches have the same ω value. The "two-ratio" model then allows one or more branches (the "foreground" branches) to have one ω value, while all other branches (the "background" branches) share another ω value. The log-likelihood score of the two-ratio model is compared with that of the one-ratio model in an LRT. Twice the difference between the two log-likelihood scores, $2\Delta\ell$, is calculated by subtracting the log-likelihood of the one-ratio model from that of the two-ratio model, and compared with a χ^2_1 distribution. One degree of freedom is used because there is one parameter difference between the two models: the one-ratio model allows a single ω value and the two-ratio model allows two. If the LRT is statistically significant, and $\omega > 1$ for the foreground branch, this is evidence for adaptive evolution. Three-ratio models allow three different values of ω in the phylogeny, and so on. The branch-based models average ω over the whole alignment, however. This means that this test will only detect positive selection so

long as the average ω value is greater than 1. If only a small number of residues are subject to even very strong positive selection, this signal will be essentially diluted by that of all the other residues which are not. Hence, selection acting on specific sites may not be detected using this method (Yang, 2006).

Site-based tests for positive selection allow a separate ω value for each codon site in the alignment (Nielsen and Yang, 1998; Yang *et al.*, 2000). Several models have been developed to do this (Anisimova *et al.*, 2001, 2002; Swanson *et al.*, 2003; Wong *et al.*, 2004). The three used in this study are known as M7, M8 and M8a. M7 and M8a are null models, which do not allow the ω of any sites to be greater than 1. M8 is the alternative model, and thus adds another class of sites with $\omega > 1$. The $2\Delta\ell$ of the M8 *vs.* M7 LRT is compared with a χ^2_2 distribution, as M8 has two more parameters than M7, though this is expected to make the test conservative (Yang, 2006). Also, it is thought that in some cases the M8 model could be a better fit to the data than M7, even if there is actually no positive selection, meaning that this test could result in false positives (Swanson *et al.*, 2003). The M8a model is essentially the same as M8 but with $\omega = 1$. There is therefore one parameter difference between the models, so in the M8 *vs.* M8a LRT, a χ^2_1 distribution is used. This test produces fewer false positives than M8 *vs.* M7. It also more powerful in some cases because of the reduction in degrees of freedom (Swanson *et al.*, 2003; Wong *et al.*, 2004). While the site-based tests allow a separate ω for each site, and are therefore better at detecting positive selection affecting only a few codons in an alignment, they instead average ω over the whole phylogeny. This means that if only certain lineages are undergoing adaptive evolution, again the signal of this selection may be drowned out by that of the other branches in the phylogeny that are not (Yang, 2006).

As mentioned earlier, positive selection is thought to most often affect a small number of codon sites in a small number of branches on a phylogeny. The tests described so far address either individual branches or individual sites, but not the combination. The branch-site models, however, are able to detect positive selection on individual lineages while at the same time allowing codon sites in the alignment to have differing ω values (Yang and Nielsen, 2002). Branches are defined by the user as

foreground or background branches, as in the branch models. The alternate model, called branch-site model A, allows codon sites to fall into one of four categories, depending on whether they are on a foreground or background branch. On both foreground and background branches, site class 0 has $0 < \omega_0 < 1$, and site class 1 has $\omega_1 = 1$. The two other site classes, 2a and 2b, differ for foreground and background branches: on foreground branches, both 2a and 2b have $\omega_2 > 1$; on background branches 2a has $0 < \omega_0 < 1$, and 2b has $\omega_1 = 1$. Model A is compared in an LRT with a null model, which constrains ω_2 to 1, thus not allowing positive selection. The $2\Delta\ell$ value is compared with a χ^2_1 distribution (Yang *et al.*, 2005; Zhang *et al.*, 2005). This test has also been shown to have low false positive rates compared with similar tests, and is capable of distinguishing between actual instances of positive selection and situations where selective constraints are relaxed (Zhang *et al.*, 2005).

All these models are implemented in the software package PAML (Phylogenetic Analysis by Maximum Likelihood) by Yang (1997). The user provides codon alignments and tree topologies, with foreground branches marked for the branch and branch-site models, and a control file of parameters to specify the model to be run. The models return the log-likelihood scores which are used to construct the LRT. For the site-based model M8 and the branch-site model A, the program also outputs posterior probabilities for sites likely to be under positive selection, estimated by Bayes empirical Bayes (BEB) analysis. In this chapter, PAML was used to apply these models to the desaturase genes identified in Chapter 2, to investigate whether their evolution has been shaped by diversifying selection. Of particular interest are duplicated genes, which in other gene families have been found to exhibit evidence relaxed constraints and in some cases positive selection, which can indicate neofunctionalisation (Gardiner *et al.*, 2008; Guo and Kim, 2007; Hahn *et al.*, 2007; Lynch and Conery, 2000). A summary of the work presented in this chapter appears in Keays *et al.* (2011).

3.2 Methods

Site-based tests were used on each gene. For duplicated genes, only the putative most ancestral gene was subjected to the site-based tests. *D. willistoni desatF* was not included in the site-based test for *desatF* because neither duplicate is orthologous with *desatFα* in the other species. For genes that have undergone duplication in some species, the branch and branch-site tests were employed. All paralogs were included in the branch and branch-site tests. This was to determine whether selection pressures differ before and after duplication, and whether there is any evidence for adaptive evolution following a duplication event.

Analyses were performed using the putative species tree (Figure 1.2), and also using subtrees of only more closely-related species. The reasoning for subtree analysis is because the species tree contains species that are in fact quite distantly-related, it may lead to wrong assumptions when used in the PAML analyses. Using subtrees of only more closely-related species attempts to minimise this and better ensure that instances of positive selection detected are "real" (Keays *et al.*, 2011).

3.2.1 Sequence alignments

PAML requires a multiple alignment of codon sequences in order to estimate levels of synonymous and nonsynonymous substitutions. Amino acid sequences for each orthologous group were aligned using CLUSTALW v2.0.12 (Thompson *et al.*, 1994). The PROTAL2DNA script v2.0 (Schuerer and Ledontal) was used to align the corresponding nucleotide sequences, based on the amino acid alignments.

3.2.2 Phylogenies

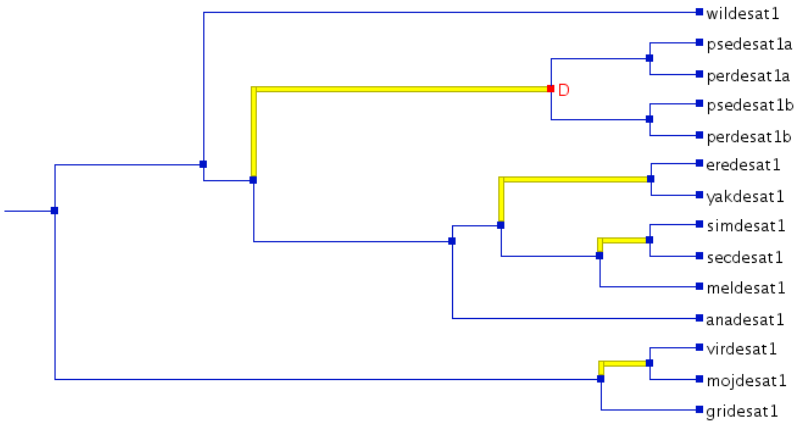
PAML also requires an unrooted phylogenetic tree showing the relationship between the sequences in the codon alignment. For site-based analyses, the putative species tree (Figure 1.2) was used, only including the most ancestral paralog of any duplicated genes. For the branch and branch-site tests on duplicated genes, maximum likelihood phylogenies were reconstructed using amino acid sequences. The appropriate substitution model was estimated using the AIC criterion in MODEL-

GENERATOR (Katoh *et al.*, 2002) with four Γ categories. The gene phylogenies were created using TREEFINDER (Jobb *et al.*, 2004), with branch support estimated using Local Rearrangements-Expected Likelihood Weights (LR-ERW) with 1000 replicates. These gene trees were then reconciled with the species tree using NOTUNG (Durand *et al.*, 2006), allowing rearrangement on branches with less than 80% LR-ERW support. Tree reconciliation compares the putative species tree with the gene tree, here created with TREEFINDER, and enables gene losses and gains to be inferred and mapped onto the species tree (Durand *et al.*, 2006; Hahn, 2007). The reconciled phylogenies are shown in Figure 3.1.

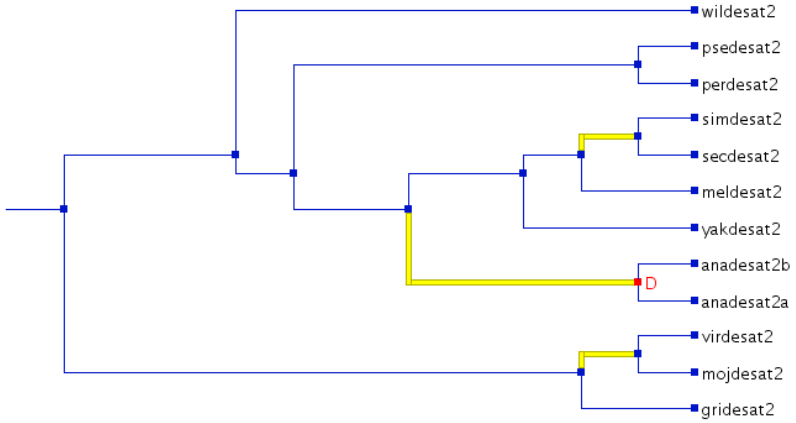
3.2.3 PAML tests

Analyses using the species tree

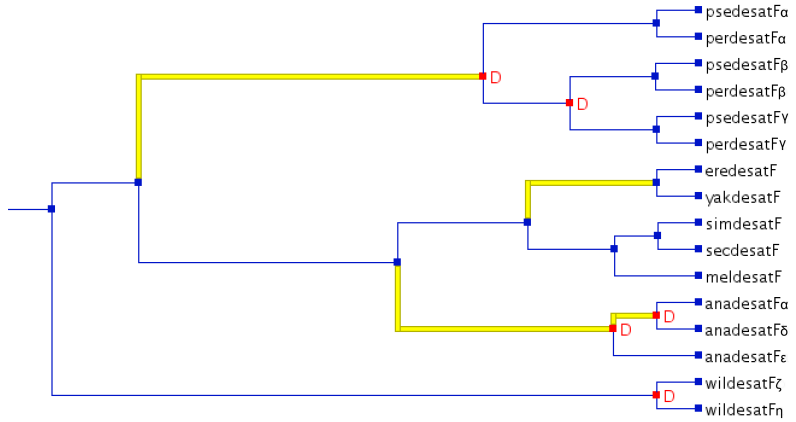
The site-based models M7, M8 and M8a were implemented for all genes. The branch and branch-site tests were only implemented for genes which had undergone duplication in one or more species: *desat1*, *desat2* and *desatF*. All foreground branches tested are shown in Figures 3.2 to 3.6. The species tree was divided into three subtrees which were used in site-based analyses with the M7, M8 and M8a models. These are shown in Figure 3.7. Subtrees were also created to analyse the duplicated genes; these are shown in Figures 3.8 to 3.12. Examples of PAML control files are given in Appendix B.



(a) Reconciled tree for *desat1*

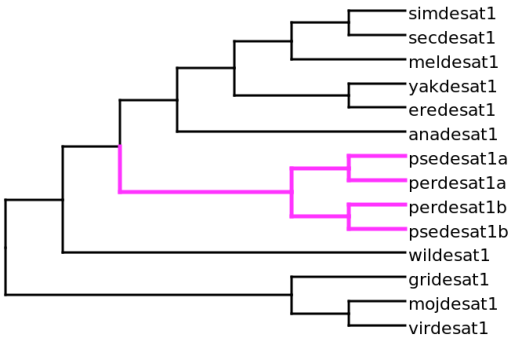


(b) Reconciled tree for *desat2*

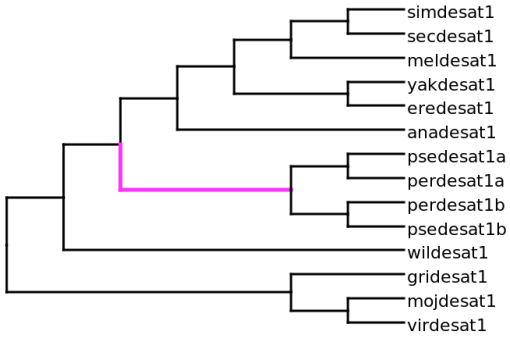


(c) Reconciled tree for *desatF*

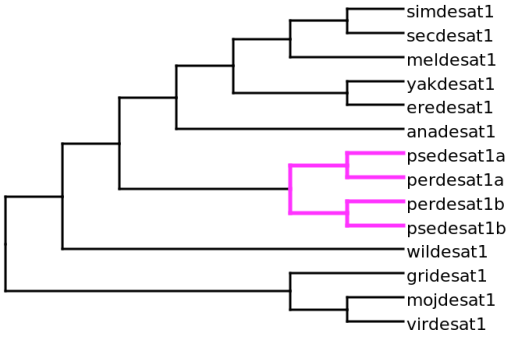
Figure 3.1: Gene trees after reconciliation with the species tree. Branches highlighted in yellow had LR-ERW support less than 80% and were thus subject to rearrangement by NOTUNG. Red squares with "D" indicate duplication events.



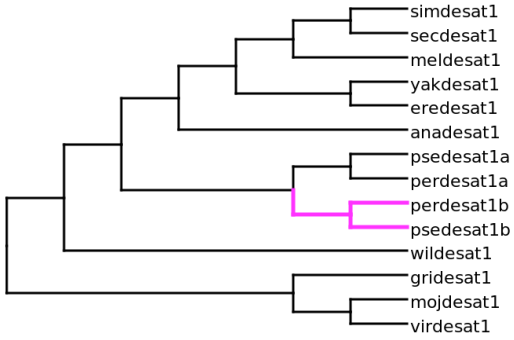
(a) *obscura* group *desat1* clade



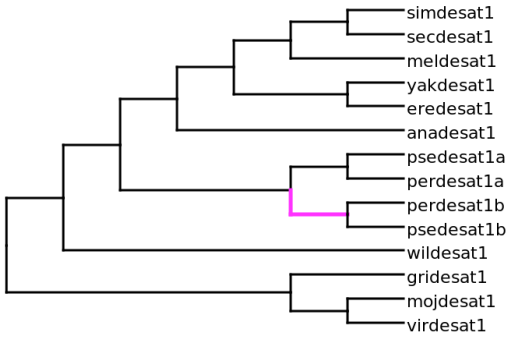
(b) *obscura* group *desat1* pre-duplication



(c) *obscura* group *desat1* post-duplication

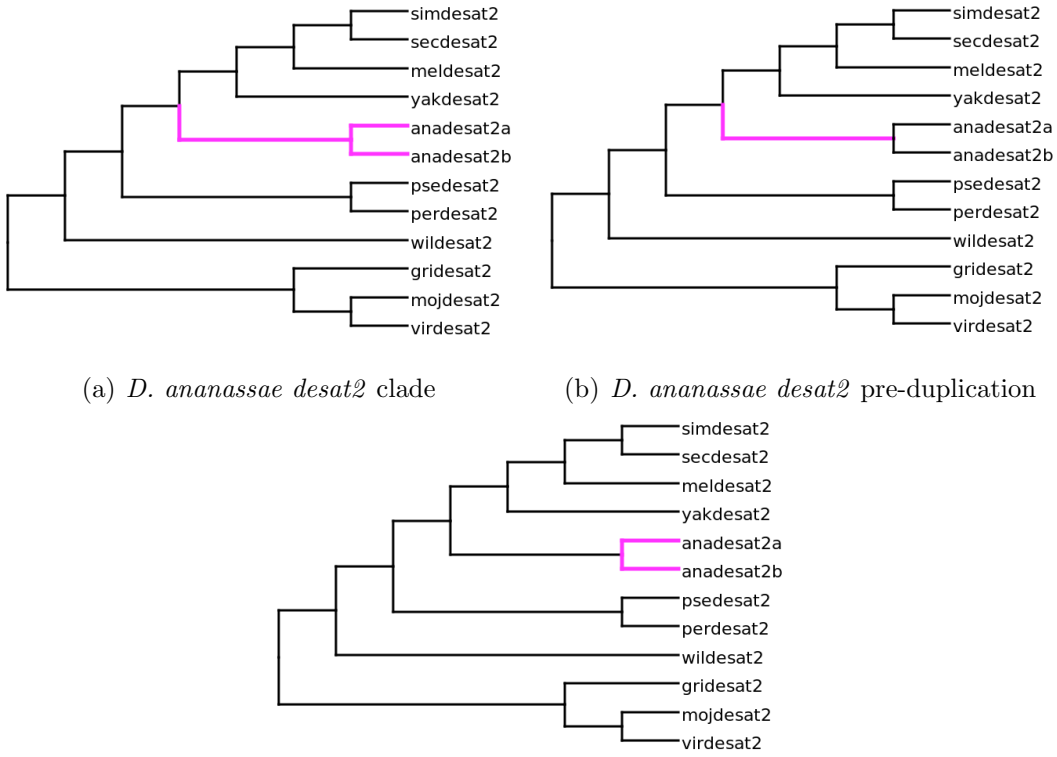


(d) *obscura* group *desat1b*



(e) Branch leading to *obscura* group *desat1b* (pre-speciation)

Figure 3.2: Trees used in branch and branch-site tests on *desat1*. Foreground branches are shown in pink; background branches are shown in black.

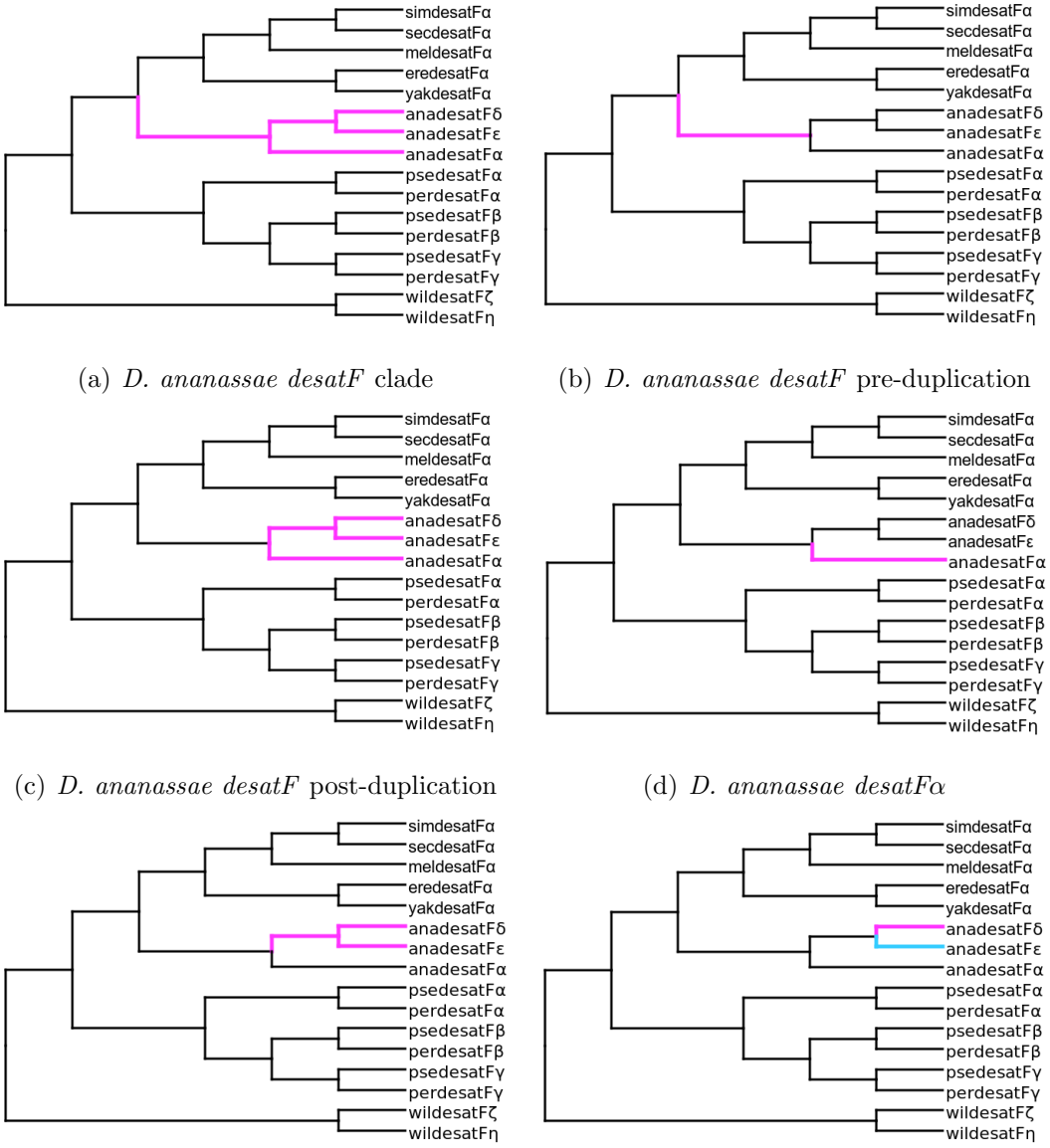


(a) *D. ananassae desat2* clade

(b) *D. ananassae desat2* pre-duplication

(c) *D. ananassae desat2* post-duplication

Figure 3.3: Trees used in branch and branch-site tests on *desat2*. Foreground branches are shown in pink; background branches are shown in black.



(a) *D. ananassae desatF* clade

(b) *D. ananassae desatF* pre-duplication

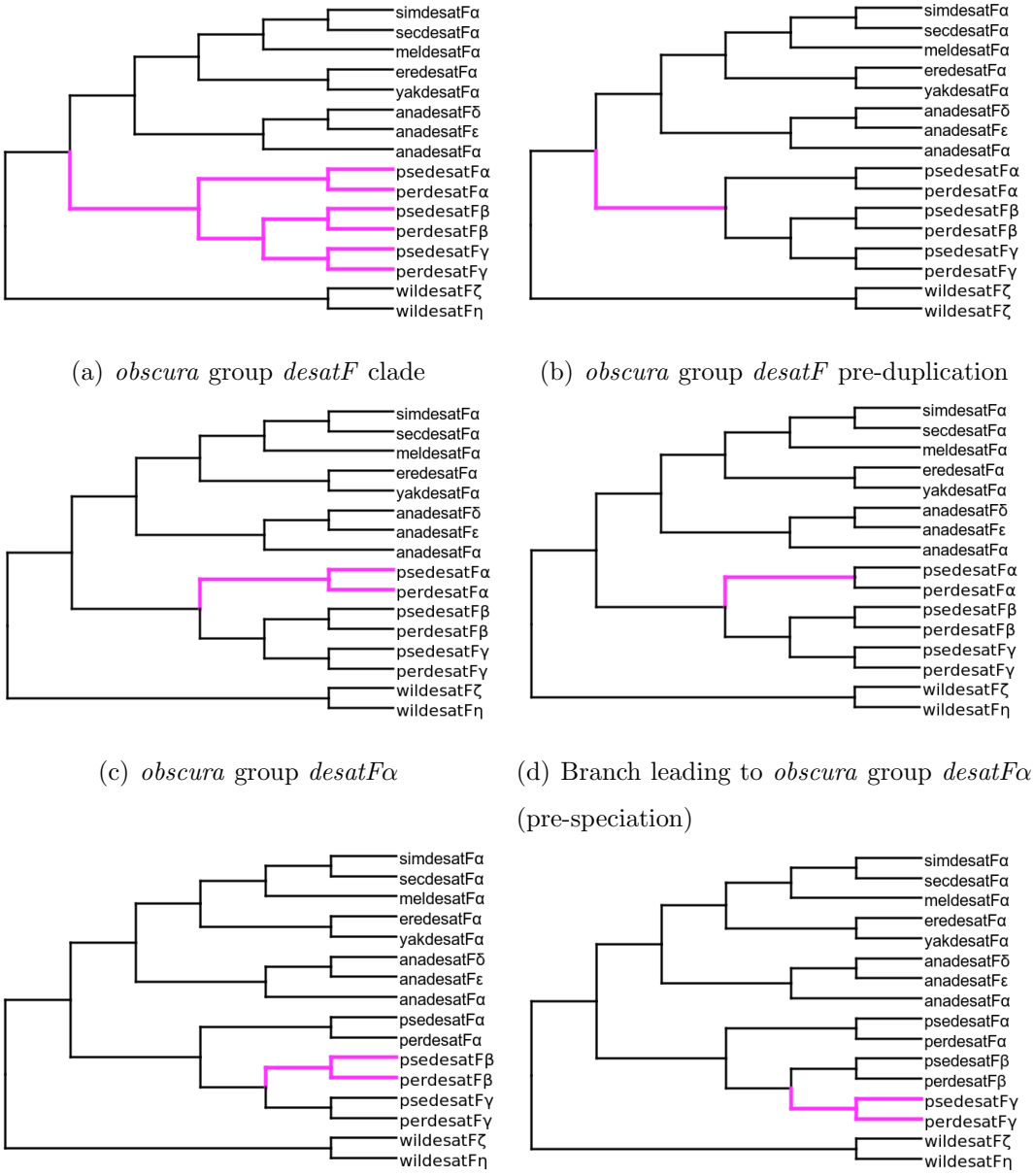
(c) *D. ananassae desatF* post-duplication

(d) *D. ananassae desatFα*

(e) *D. ananassae desatFδ* and *desatFϵ* 2-ratio model

(f) *D. ananassae desatFδ* and *desatFϵ* 3-ratio model

Figure 3.4: Trees used in branch and branch-site tests on *D. ananassae desatF*. Foreground branches are shown in pink and blue to denote differing ω estimates; background branches are shown in black.



(a) *obscura* group *desatF* clade

(b) *obscura* group *desatF* pre-duplication

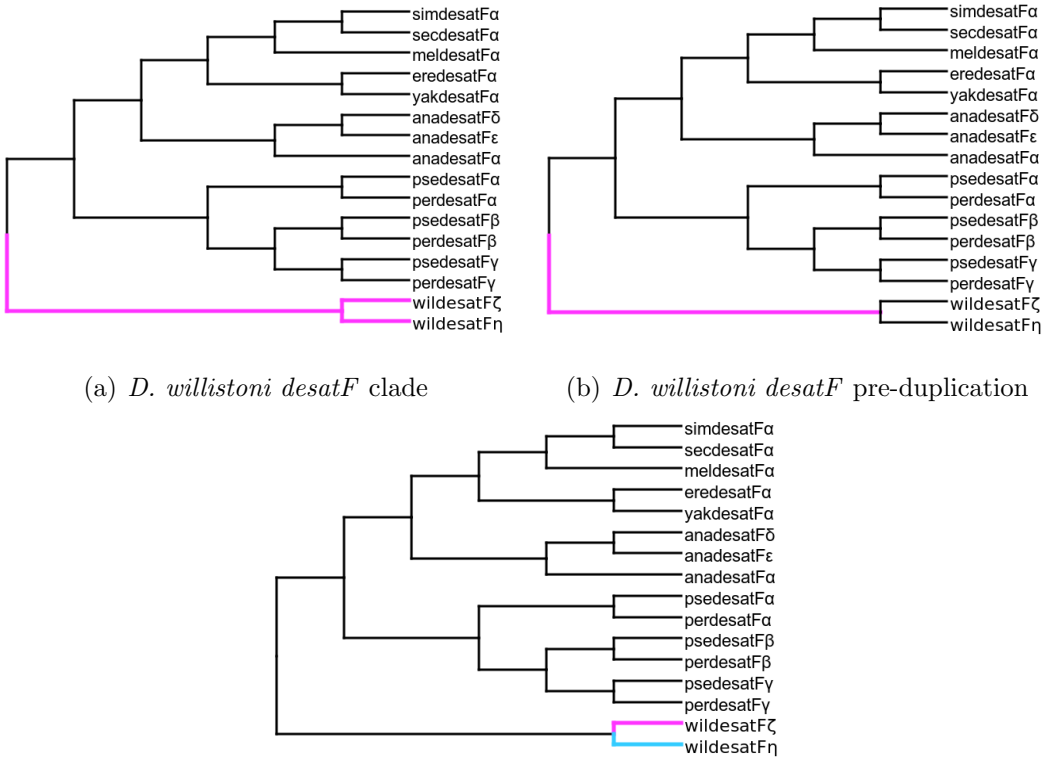
(c) *obscura* group *desatFα*

(d) Branch leading to *obscura* group *desatFα* (pre-speciation)

(e) *obscura* group *desatFβ*

(f) *obscura* group *desatFγ*

Figure 3.5: Trees used in branch and branch-site tests on *obscura* group *desatF*. Foreground branches are shown in pink; background branches are shown in black.

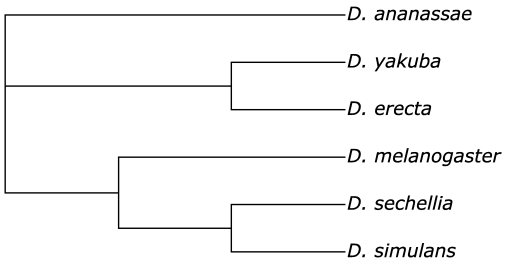


(a) *D. willistoni desatF* clade

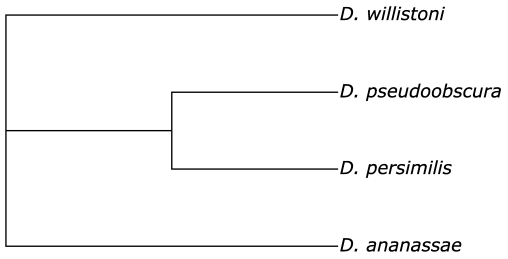
(b) *D. willistoni desatF* pre-duplication

(c) *D. willistoni desatFζ* and *desatFη* 3-ratio model

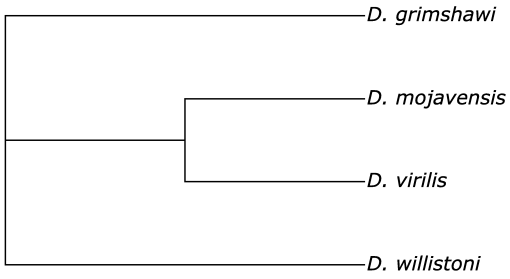
Figure 3.6: Trees used in branch and branch-site tests on *D. willistoni desatF*. Foregound branches are shown in pink and blue; background branches are shown in black.



(a) *melanogaster* group

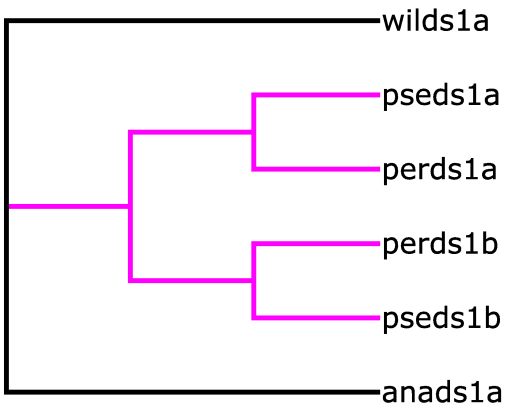


(b) *D. ananassae*, *obscura* group and *D. willistoni*

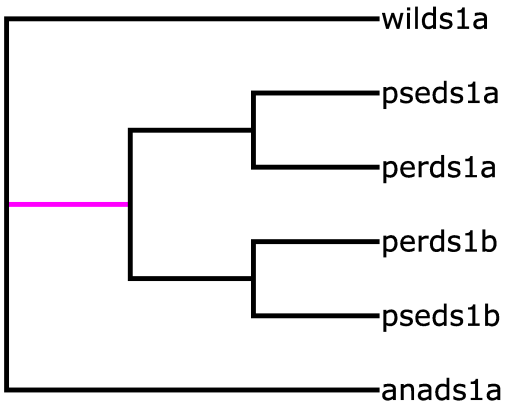


(c) *D. willistoni*, *D. mojavensis* and *D. grimshawi*

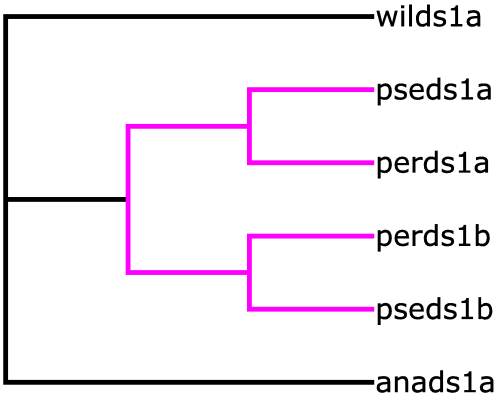
Figure 3.7: Subtrees used in site-based PAML analyses.



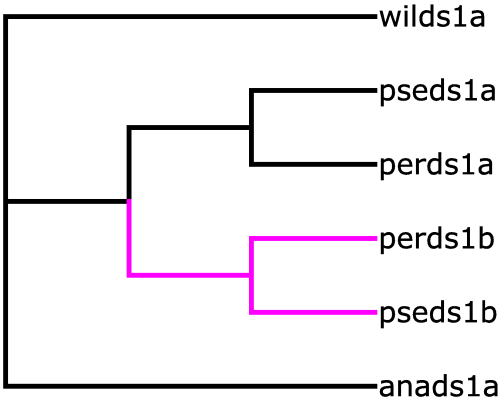
(a) *obscura* group *desat1* clade



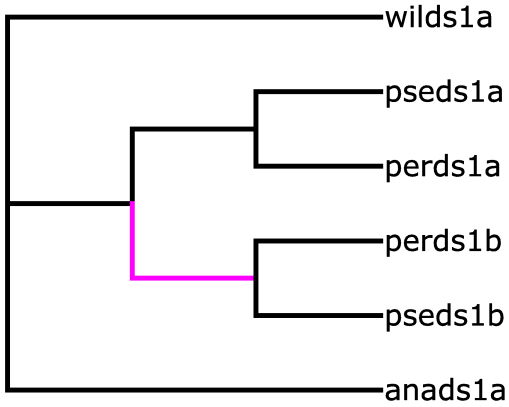
(b) *obscura* group *desat1* pre-dupication



(c) *obscura* group *desat1* post-duplication

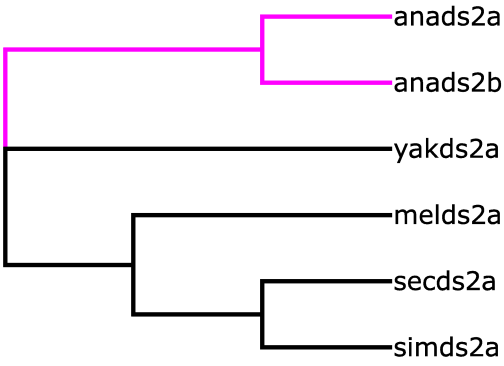


(d) *obscura* group *desat1b*

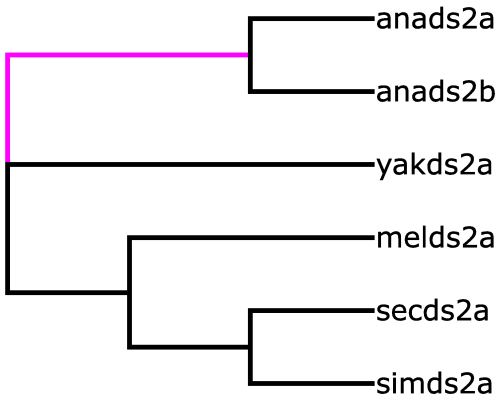


(e) Branch leading to *obscura* group *desat1b* (pre-speciation)

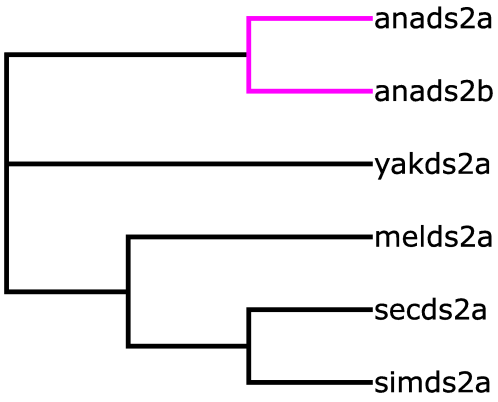
Figure 3.8: Subtrees used in branch and branch-site analyses of *obscura* group *desat1*. Foreground branches are shown in pink; background branches are shown in black.



(a) *D. ananassae desat2* clade

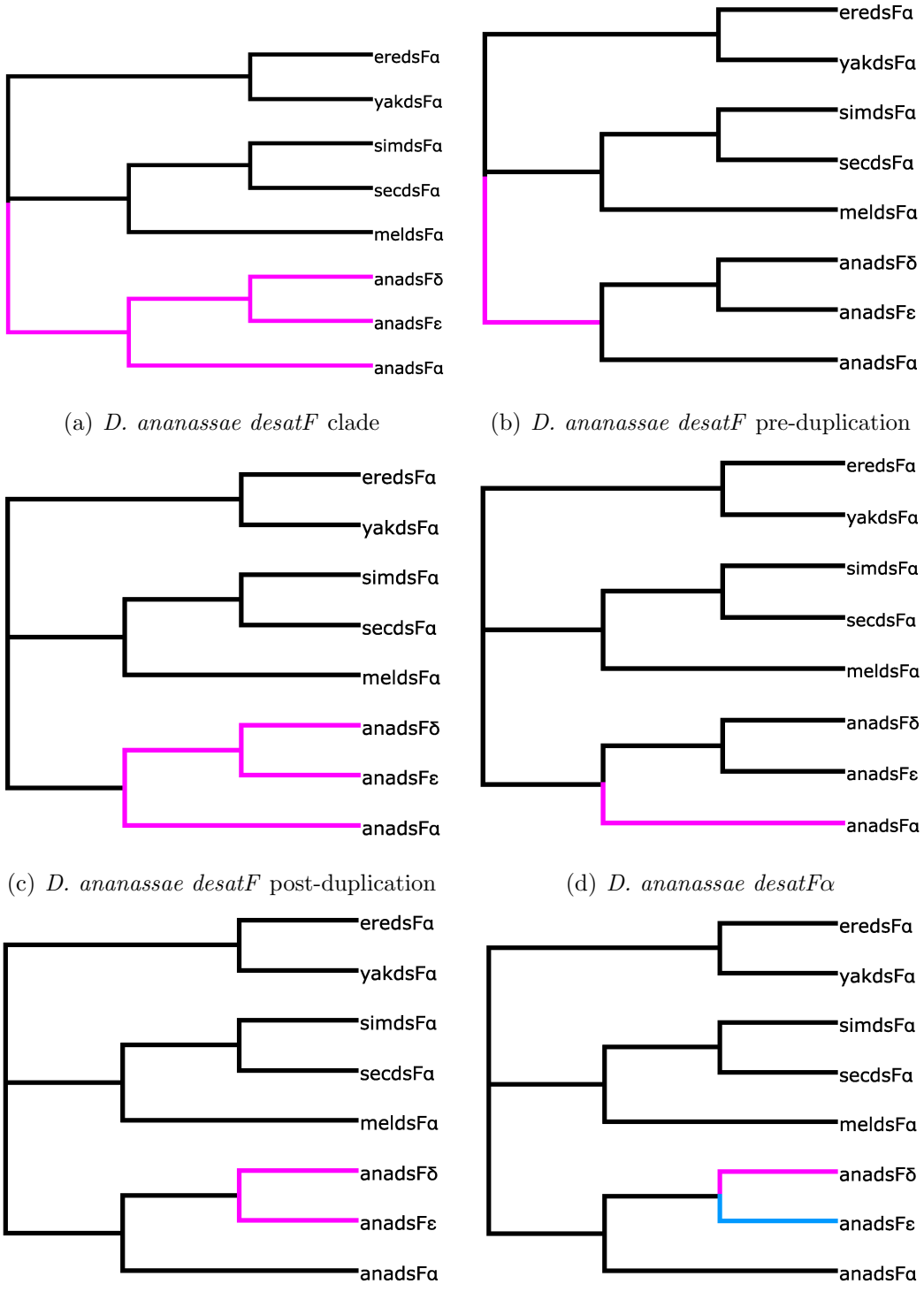


(b) *D. ananassae desat2* pre-duplication



(c) *D. ananassae desat2* post-duplication

Figure 3.9: Subtrees used in branch and branch-site analyses of *D. ananassae desat2*. Foreground branches are shown in pink; background branches are shown in black.



(a) *D. ananassae desatF* clade

(b) *D. ananassae desatF* pre-duplication

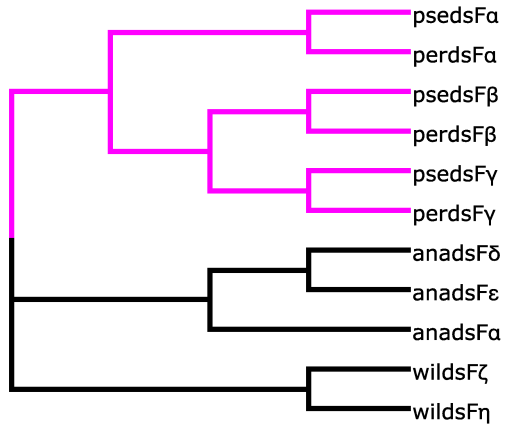
(c) *D. ananassae desatF* post-duplication

(d) *D. ananassae desatFα*

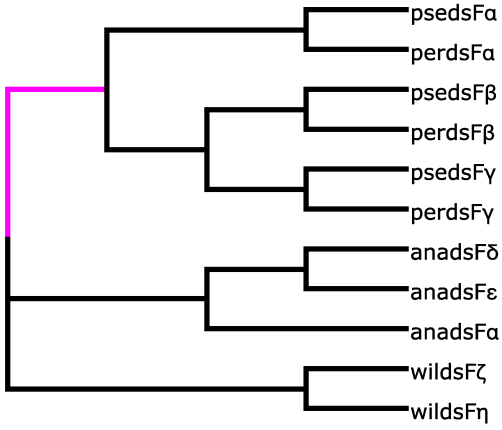
(e) *D. ananassae desatFδ* and *desatFϵ* 2-ratio

(f) *D. ananassae desatFδ* and *desatFϵ* 3-ratio

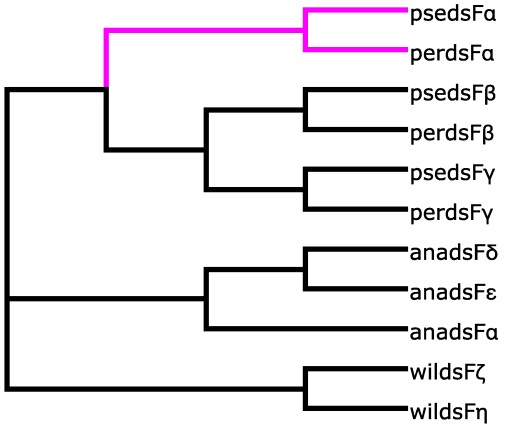
Figure 3.10: Subtrees used in branch and branch-site analyses of *D. ananassae desatF*. Foreground branches are shown in pink and blue to denote differing ω estimates; background branches are shown in black.



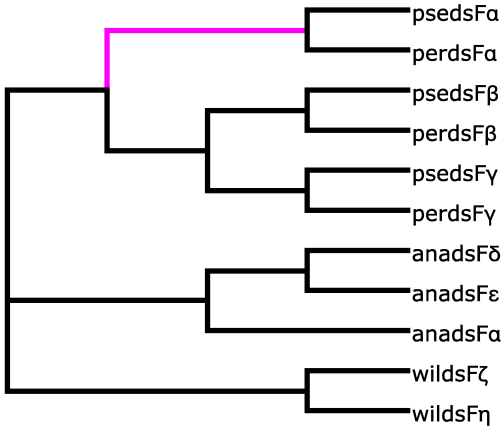
(a) *obscura* group *desatF* clade



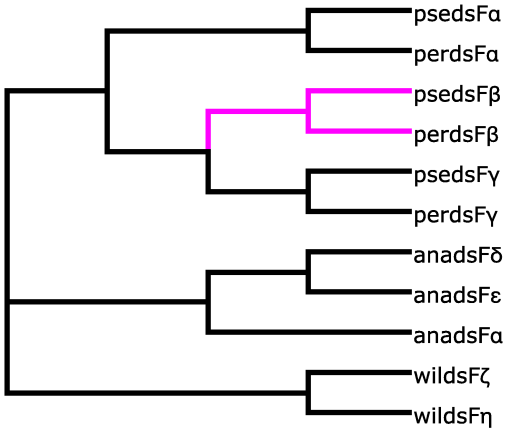
(b) *obscura* group *desatF* pre-duplication



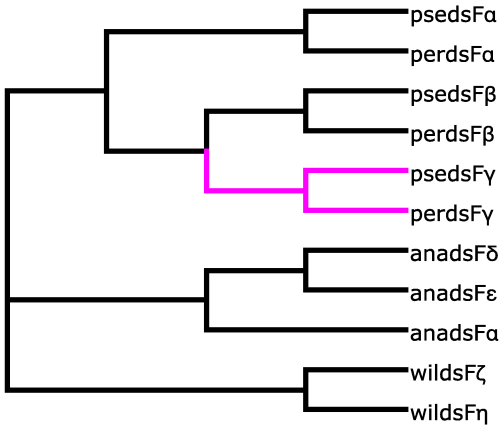
(c) *obscura* group *desatFα*



(d) Branch leading to *obscura* group *desatFα* (pre-speciation)

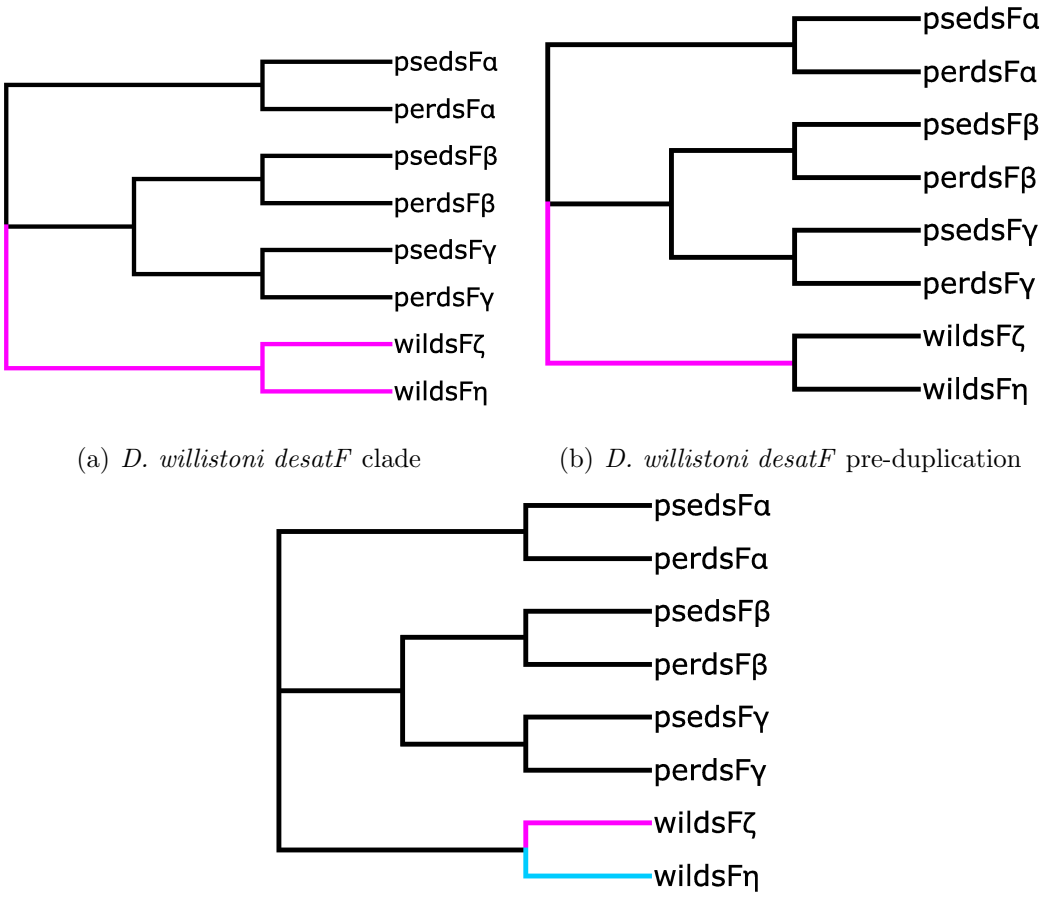


(e) *obscura* group *desatFβ*



(f) *obscura* group *desatFγ*

Figure 3.11: Subtrees used in branch and branch-site analyses of *obscura* group *desatF*. Foreground branches are shown in pink; backgroun branches are shown in black.



(a) *D. willistoni desatF* clade

(b) *D. willistoni desatF* pre-duplication

(c) *D. willistoni desatF* post-duplication

Figure 3.12: Subtrees used in branch and branch-site analyses of *D. willistoni desatF*. Foreground branches are shown in pink and blue to denote differing ω estimates; background branches are shown in black.

3.3 Results

3.3.1 Site-based analyses

No evidence of positive selection was detected with site-based analyses of the whole species tree (Table 3.1). For the subtree analyses, some of the site-based tests were significant with P<0.05 (Table 3.2). The M8 model lists posterior probabilities for codon sites under positive selection. For those genes where the LRT was significant with P<0.05, all the posterior probabilities for codons being under positive selection were nonsignificant (less than 0.949). Hence no positive selection was detected.

Gene	$2\Delta\ell$ (M7 *vs.* M8)	P value (M7 *vs.* M8)	$2\Delta\ell$ (M8a *vs.* M8)	P value (M8a *vs.* M8)
CG15531	-0.0006	1.0000	0.0014	0.9702
CG17928	3.7522	0.1532	0.9284	0.3353
CG8630	1.2575	0.5333	0.6069	0.4360
CG9743	0.1050	0.5755	0.0000	1.0000
CG9747	-0.0002	1.0000	0.0015	0.9691
Cyt-b5-r	1.2111	0.5458	-0.1097	1.0000
desat1	2.2082	0.3315	0.6614	0.4200
desat2	2.3794	0.3043	1.7120	0.1907
desatF	0.4237	0.8091	0.4246	0.5147

Table 3.1: Results of site-based tests using the whole species tree. Negative $2\Delta\ell$ values treated as 0. All P values are nonsignificant ($\alpha = 0.05$).

Gene	Tree	$2\Delta\ell$ (M7 *vs.* M8)	P value (M7 *vs.* M8)	$2\Delta\ell$ (M8a *vs.* M8)	P value (M8a *vs.* M8)
CG15531	*melanogaster* group	-0.000452	1.0000	0.000000	1.0000
	Mid section	10.86332	*0.0044*	7.317958	*0.0068*
	Lower section	0.015471	0.9923	0.006950	0.9336
CG17928	*melanogaster* group	0.000972	0.9995	-0.001856	1.0000
	Mid section	-0.001806	1.0000	0.000000	1.0000
	Lower section	6.369314	*0.0414*	2.921360	0.0874
CG8630	*melanogaster* group	-0.000316	1.0000	0.000000	1.0000
	Mid section	6.975244	*0.0306*	4.364722	*0.0367*
	Lower section	-0.000156	1.0000	0.000000	1.0000
CG9743	*melanogaster* group	-0.000588	1.0000	-0.000190	1.0000
	Mid section	-0.000034	1.0000	0.000000	1.0000
	Lower section	6.670512	*0.0356*	5.717572	*0.0168*
CG9747	*melanogaster* group	-0.001240	1.0000	0.000000	1.0000
	Mid section	0.982076	0.6120	0.158754	0.6903
	Lower section	0.267632	0.8748	0.028408	0.8662
Cyt-b5-r	*melanogaster* group	0.757590	0.6847	0.000000	1.0000
	Mid section	-0.000520	1.0000	0.000330	0.9855
	Lower section	2.482048	0.2891	1.363178	0.2430
desat1	*melanogaster* group	0.208344	0.9011	0.000000	1.0000
	Mid section	0.961706	0.6183	0.000000	1.0000
	Lower section	0.300870	0.8603	0.011894	0.9132
desat2	*melanogaster* group	7.644446	*0.0219*	3.348258	0.0673
	Mid section	-0.000360	1.0000	0.000000	1.0000
	Lower section	2.677108	0.2622	0.039310	0.8428

Table 3.2: Results of site-based tests using subtrees. Negative $2\Delta\ell$ values treated as 0. P-values less than 0.05 in italics; No P-values significant after Bonferroni correction (24 tests; P<0.002).

3.3.2 Branch and branch-site tests

Analyses using the whole species tree

Results from the branch and branch-site analyses using the whole species tree are shown in Tables 3.3 and 3.4. Eight of the 23 LRTs on 2-ratio branch-based models *vs.* the 1-ratio model were significant after Bonferroni correction (P<0.002), but all foreground ω values were much less than 1 (Table 3.3). Three of the branch-site tests revealed significant evidence of positive selection after Bonferroni correction (7 tests; P<0.007) (Table 3.4). In the analysis of *D. persimilis* and *D. pseudoobscura desat1*, the branch test was significant in the branch leading to the *desat1b* genes (Figure 3.8(e), P=2.9×10^{-6}), though the ω value of this branch was estimated at 0.15. The branch-site test on the same branch was also significant (P = 1.809×10^{-5}), and indicates that 13.7% of codon sites are under strong positive selection with ω_2=8.12. The majority of the remaining sites (80.0%) are under strong purifying selection, with ω_0=0.03.

In the analysis of *D. persimilis* and *D. pseudoobscura desatF*, the branch-site test on the branch leading to the *desatFα* genes was significant ($P = 1.007 \times 10^{-4}$). The model found ω_2=998.94. Very high estimates can arise if the value of dS is 0, and therefore ω can be taken as infinity (Yang, 2005). A 95% confidence interval for ω_2 was constructed by running the branch-site model A with ω_2 fixed at each integer in the range $0 \leq \omega_2 \leq 2000$. ℓ from each run, and ℓ from the original model in which ω_2 was estimated by maximum likelihood as ω_2=998.94, were compared using the LRT given in equation 3.1.

$$2\Delta\ell = 2(\ell_{\omega_2=998.94} - \ell_{\omega_2 fixed}) \tag{3.1}$$

Figure 3.13 shows that a broad range of ω_2 values have the maximum ℓ of -8288.727, including $\omega_2 = 998.94$. Figure 3.14 shows the P-values from the LRTs (1 d.f.). When P=1, $2\Delta\ell = 0$, i.e. the two ℓ values are the same. The 95% confidence interval for ω_2 comprises all values for which P>0.05, as indicated by the red line in Figure 3.14. The lowest value above this line is $\omega_2 = 5$. ω_2 therefore has an estimated confidence interval with a lower limit of 5, and an upper limit which

exceeds 2000, the highest value used here.

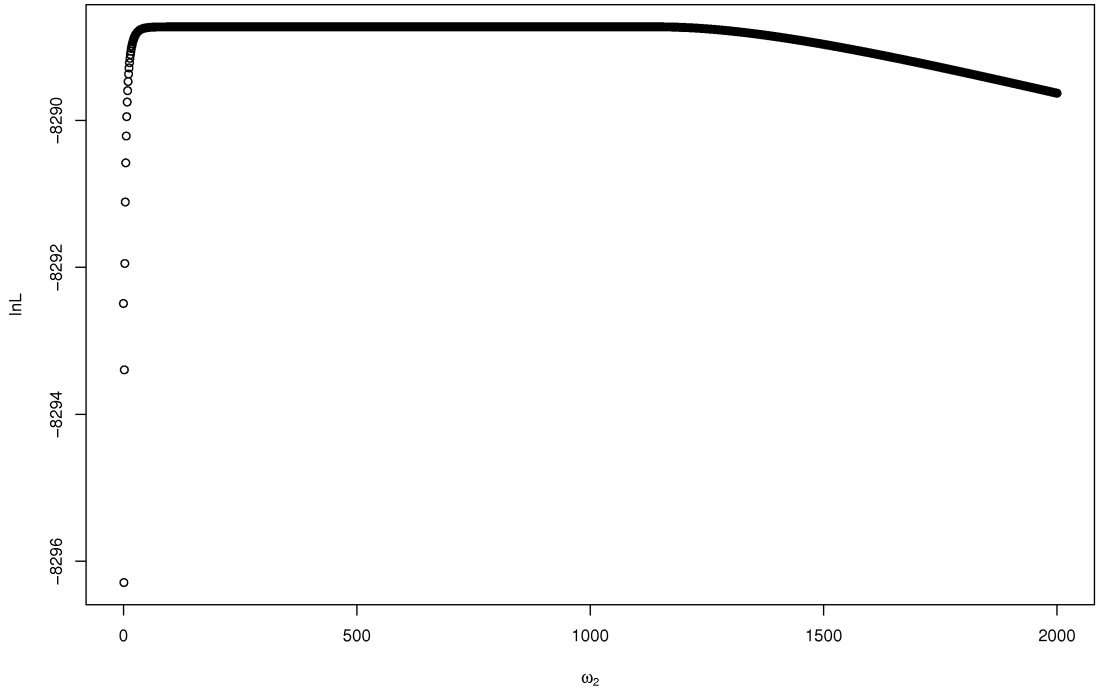


Figure 3.13: Graph of ℓ *vs.* ω_2 for *obscura* group *desatF*α.

Gene	Species	Branch(es) tested	$2\Delta\ell$	P value	Foreground ω
desat1	*obscura* group	Figure 3.2(a)	19.64	$\mathbf{9.35\times10^{-6}}$	0.11
		Figure 3.2(b)	1.40	0.24	0.17
		Figure 3.2(c)	19.36	$\mathbf{1.08\times10^{-5}}$	0.12
		Figure 3.2(d)	19.62	$\mathbf{9.45\times10^{-6}}$	0.14
		Figure 3.2(e)	21.88	$\mathbf{2.9\times10^{-6}}$	0.15
desat2	*D. ananassae*	Figure 3.3(a)	0.42	0.52	0.07
		Figure 3.3(b)	3.8×10^{-3}	0.95	0.08
		Figure 3.3(c)	0.47	0.50	0.07
desatF	*D. ananassae*	Figure 3.4(a)	14.53	$\mathbf{10^{-4}}$	0.08
		Figure 3.4(b)	3.5×10^{-4}	0.99	0.12
		Figure 3.4(c)	15.87	$\mathbf{6.7\times10^{-5}}$	0.07
		Figure 3.4(d)	40.33	$\mathbf{2.14\times10^{-10}}$	0.02
		Figure 3.4(e)	0.33	0.57	0.11
		Figure 3.4(f)	2.77	0.1	ω_δ=0.14; ω_ϵ=0.08
	obscura group	Figure 3.5(a)	0.40	0.53	0.12
		Figure 3.5(b)	3.07	0.08	0.02
		Figure 3.5(c)	4.11	*0.04*	0.07
		Figure 3.5(d)	2.15	0.14	0.08
		Figure 3.5(e)	5.09	*0.02*	0.24
		Figure 3.5(f)	11.85	$\mathbf{5.8\times10^{-4}}$	0.28
	D. willistoni	Figure 3.6(a)	2.54	0.11	0.16
		Figure 3.6(b)	0.00	1.00	0.85
		Figure 3.6(c)	3.61	0.06	ω_ς=0.1; ω_η=0.2

Table 3.3: Results of branch tests using whole species tree. P-values less than 0.05 in italics; P-values significant after Bonferroni correction (23 tests; P<0.002) in bold. All ω values are less than 1.

Gene	Species	Branch(es) tested	$2\Delta\ell$	P value	ω_2
desat1	*obscura* group	Figure 3.2(e)	18.38	$\mathbf{1.81\times10^{-5}}$	**8.12**
desatF	*D. ananassae*	Figure 3.4(d)	3.58	0.06	369.16
		Figure 3.4(e)	0.00	1.00	1.00
	obscura group	Figure 3.5(c)	9.58	$\mathbf{1.97\times10^{-3}}$	**5.52**
		Figure 3.5(d)	15.12	$\mathbf{1.01\times10^{-4}}$	**998.94** (∞)
		Figure 3.5(e)	0.81	0.37	3.72
		Figure 3.5(f)	0.00	1.00	1.00

Table 3.4: Results of branch-site tests using whole species tree. P-values significant after Bonferroni correction (7 tests; P<0.007) and ω_2 >1 in bold.

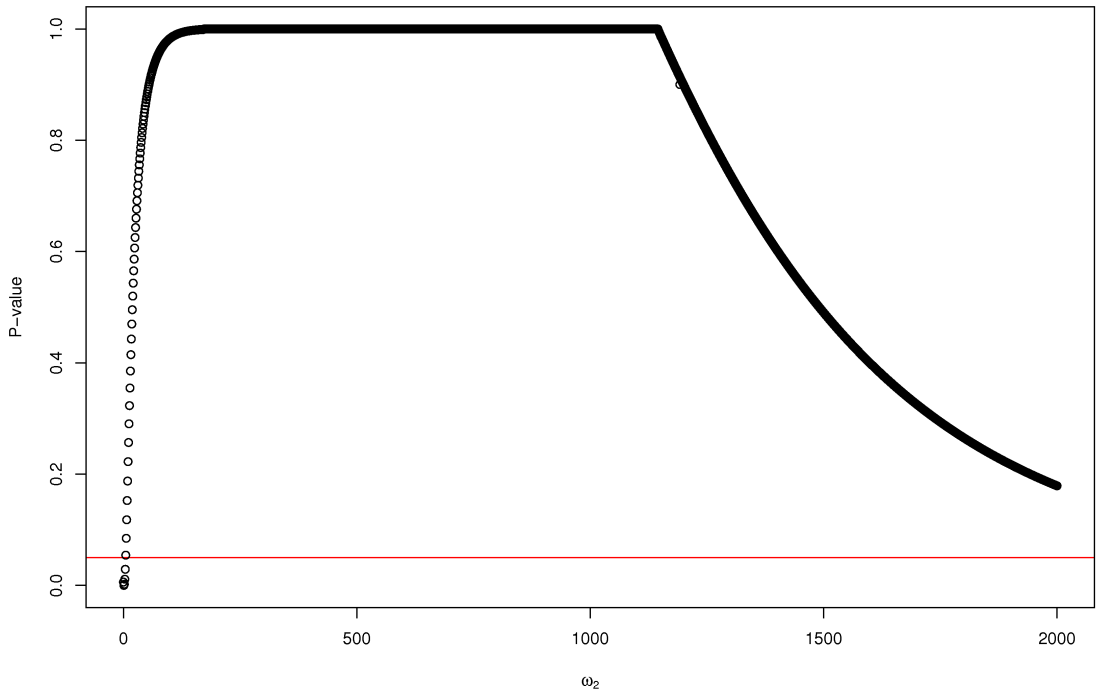


Figure 3.14: Graph of P values *vs.* ω_2 for *obscura* group *desatF*α.

Analyses using subtrees

Positive selection was detected by branch and branch-site subtree analysis in some duplicated genes. Eight of the 23 LRTs on 2-ratio branch-based models *vs.* the 1-ratio model were significant after Bonferroni correction ($P<0.002$), but all foreground ω values were much less than 1 (Table 3.5). Three of the branch-site tests revealed significant evidence of positive selection before Bonferroni correction ($P<0.05$; Table 3.6) After Bonferroni correction only one remained significant (7 tests; $P<0.007$). In the analysis of *D. persimilis* and *D. pseudoobscura desat1*, the branch test was significant in the branch leading to the *desat1b* genes (Figure 3.8(e); $P=8.97\times10^{-6}$), though the ω value of this branch was estimated at 0.14. The branch-site test on the same branch was significant before Bonferroni correction ($P=0.03$), and indicates that 11.5% of codon sites are under strong positive selection with $\omega_2=4.16$. The majority of the remaining sites (85.1%) are under strong purifying selection, with $\omega_0=0.04$. The Bayes empirical Bayes results of the alternative model show that most of the sites predicted to be under positive selection are concentrated near the N-terminus of the protein. The structure of the Desat1b protein is not currently known. If we apply the proposed secondary structure of a desaturase from the fungus *Claviceps purpurea* (Meesapyodsuk *et al.*, 2007), sites affected by positive selection, and the putative active site and transmembrane helices, are illustrated in Figure 3.15.

Gene	Species	Branch(es) tested	$2\Delta\ell$	P value	Foreground ω
desat1	*obscura* group	Figure 3.8(a)	18.24	$\mathbf{1.95 \times 10^{-5}}$	0.10
		Figure 3.8(b)	0.12	0.73	0.05
		Figure 3.8(c)	19.14	$\mathbf{1.22 \times 10^{-5}}$	0.11
		Figure 3.8(d)	17.65	$\mathbf{2.66 \times 10^{-5}}$	0.13
		Figure 3.8(e)	19.72	$\mathbf{8.97 \times 10^{-6}}$	0.14
desat2	*D. ananassae*	Figure 3.9(a)	4.06	0.04	0.06
		Figure 3.9(b)	0.97	0.33	0.06
		Figure 3.9(c)	2.80	0.09	0.06
desatF	*D. ananassae*	Figure 3.10(a)	17.38	$\mathbf{3.06 \times 10^{-5}}$	0.08
		Figure 3.10(b)	0.19	0.66	0.10
		Figure 3.10(c)	17.64	$\mathbf{2.67 \times 10^{-5}}$	0.07
		Figure 3.10(d)	39.95	$\mathbf{2.61 \times 10^{-5}}$	0.02
		Figure 3.10(e)	0.15	0.70	0.11
		Figure 3.10(f)	1.40	0.24	ω_δ=0.14; ω_ϵ=0.09
	obscura group	Figure 3.11(a)	4.69	*0.03*	0.13
		Figure 3.11(b)	1.30	0.25	0.02
		Figure 3.11(c)	2.72	0.10	0.07
		Figure 3.11(d)	0.99	0.32	0.08
		Figure 3.11(e)	7.97	$\mathbf{5.0 \times 10^{-3}}$	0.26
		Figure 3.11(f)	16.89	$\mathbf{3.95 \times 10^{-5}}$	0.29
	D. willistoni	Figure 3.12(a)	0.40	0.53	0.13
		Figure 3.12(b)	28.62	$\mathbf{8.82 \times 10^{-8}}$	0.01
		Figure 3.12(c)	9.12	$\mathbf{2.25 \times 10^{-3}}$	ω_ζ=0.10; ω_η=0.23

Table 3.5: Results of branch tests using subtrees. P-values less than 0.05 in italics; P-values significant after Bonferroni correction (23 tests; P<0.002) in bold. All ω values are less than 1.

Gene	Species	Branch(es) tested	$2\Delta\ell$	P value	Foreground ω
desat1	*obscura* group	Figure 3.8(e)	4.55	*0.03*	**4.16**
desatF	*D. ananassae*	Figure 3.10(d)	0.00	1.00	1.00
		Figure 3.10(e)	0.00	1.00	1.00
	obscura group	Figure 3.11(c)	6.27	*0.01*	**4.03**
		Figure 3.11(d)	11.01	$\mathbf{9.06 \times 10^{-4}}$	**941.43**
		Figure 3.11(e)	0.11	0.74	1.99
		Figure 3.11(f)	0.00	1.00	1.00

Table 3.6: Results of branch-site tests using subtrees. P-values less than 0.05 in italics; P-values significant after Bonferroni correction (7 tests; P<0.007) and $\omega_2 > 1$ in bold.

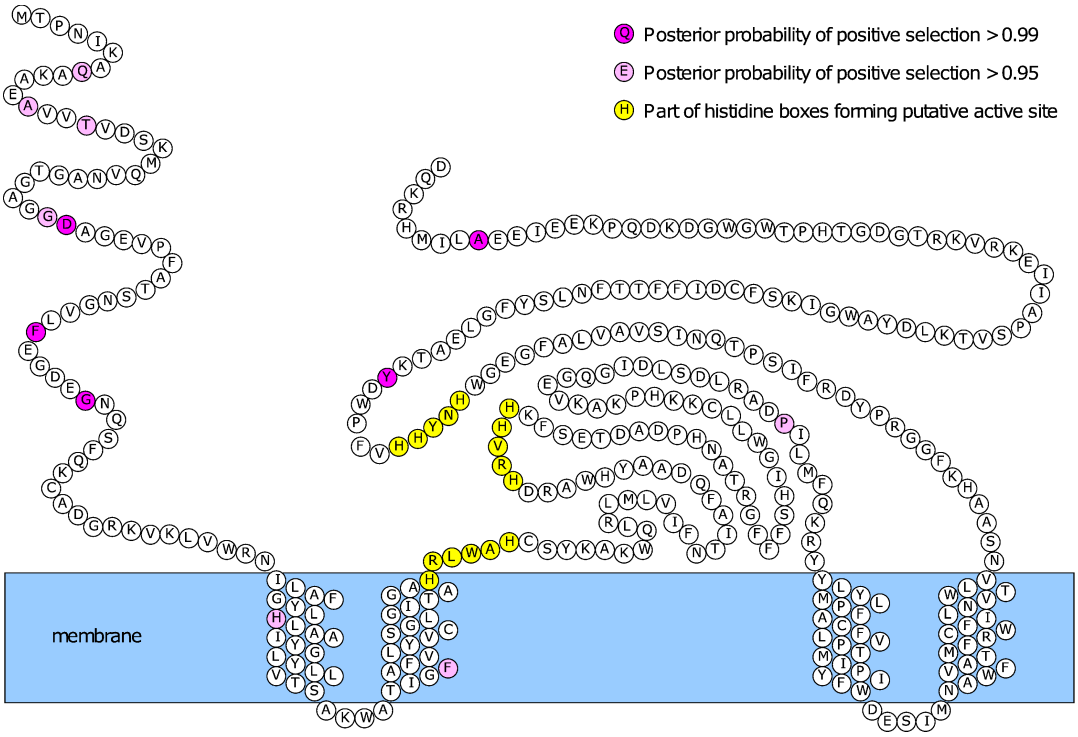


Figure 3.15: Hypothetical structure of Desat1b showing sites affected by positive selection.

3.4 Discussion

Positive selection has been found to be important in the evolution of various putative speciation genes (Orr, 2005; Wu and Ting, 2004). None of the single-copy desaturase genes showed any significant evidence for positive selection, indicating that these loci are under strong selective constraints. The results from the branch tests on duplicated genes are consistent with this hypothesis: even where an LRT was significant, all ω values were much less than 1, suggesting that on the whole the loci are under negative selection. The lack of significant LRTs, and the low ω values in the branch tests, may be artefacts of the way the models average ω over the entire phylogeny or alignment. This would mean selection on individual branches or codon sites is not picked up (Yang, 2006). However, it still suggests that overall, these genes are under purifying selection. Apart from acting as pheromones, cuticular hydrocarbons are important for *Drosophila* physiology, having vital roles in temperature tolerance and desiccation resistance (Rouault *et al.*, 2000). It is therefore perhaps not surprising that the genes involved in their production seem to be strongly constrained by selection.

The branch-site tests have detected some instances of positive selection. In the analysis of *obscura* group *desatF*, selection was detected in one of the duplicates, *desatF*α. The selection appears to be strongest on the branch leading to the *D. pseudoobscura* and *D. persimilis* genes, just before speciation (Figures 3.5(d) and 3.11(d)). This branch has an extremely high ω_2 value (998.94 for the species tree, 941.43 for the subtree), which may not be accurate (Yang, 2005). However the LRT is still significant, so it is likely that positive selection has affected the evolution of this duplicate in these two species, suggesting possible neofunctionalisation.

Positive selection was also detected in *D. pseudoobscura* and *D. persimilis desat1b*, again on the branch preceding speciation. In this case, the majority of the positively selected sites are concentrated at the N-terminus of the protein (Figure 3.15), which is highly divergent compared with the ancestral Desat1 protein (Figure 2.4). The function of this new duplicate is not yet known, but the fact that it is closely related to a known pheromone gene showing strong evidence for positive selection makes it an interesting candidate for further study. The expression pattern

of this gene is investigated further in Chapter 4.

There were slight differences in the results from the tests using the whole species tree and those using subtrees. Most of the branch and branch-site tests that had significant P-values when the species tree was used were also significant in the subtree analysis, although some of them were not significant after Bonferroni correction. For the site-based tests, the analyses using the species tree did not produce any significant LRTs. However, in the site-based subtree analyses, seven of the LRTs had P values less than 0.05. This could be due to the way the site-based models average ω over the phylogeny: if the tree includes a branch which has some positive selection, perhaps the more other branches there are on the tree, the more the signal of this selection will be drowned out. So maybe using a smaller tree would prevent this from happening, which could be why the subtrees have more significant LRTs. However, none of the results from the BEB analysis in the M8 model showed any sites with significantly high ω values. Conversely, the branch and branch-site tests using the subtrees showed less significant results that those using the species tree. This may also be an artefact of reducing the number of branches in the subtree analyses.

The results of the analyses in this chapter show that the desaturase loci are generally experiencing strong negative selection pressure. However, the significant branch-site test results suggest that, following a duplication event, it is possible that these constraints can be relaxed on one of the paralogs, and adaptive diversifying evolution and neofuncionalisation can then occur. This has been found in other studies of gene family evolution across these species of *Drosophila* (Gardiner *et al.*, 2008; Guo and Kim, 2007; Hahn *et al.*, 2007; Vieira *et al.*, 2007). While this analysis found positive selection in the coding regions of desaturase loci, it is unable to pick up selection affecting non-coding regions. Selection on regulatory regions has been found to have effects on gene expression, for example in the *desat2* locus, where the Cosmopolitan and African populations have differing promoters which causes a difference in expression (Takahashi *et al.*, 2001).

Bibliography

Anisimova, M., Bielawski, J. P. and Yang, Z. (2001). The accuracy and power of likelihood ratio tests to detect positive selection at amino acid sites. *Molecular Biology and Evolution*, **18**:1585–1592.

Anisimova, M., Bielawski, J. P. and Yang, Z. (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution*, **19**:950–958.

Durand, D., Halldorsson, B. V. and Vernot, B. (2006). A hybrid micro-macroevolutionary approach to gene tree reconstruction. *Journal of Computational Biology*, **13**:320–335.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C. and Ritchie, M. G. (2008). *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Molecular Ecology*, **17**:1648–1657.

Guo, S. and Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular Biology and Evolution*, **24**:1198–1207.

Hahn, M. W. (2007). Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. *Genome Biology*, **8**:R141.

Hahn, M. W., Han, M. V. and Han, S. G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, **3**:e197.

Jobb, G., von Haeseler, A. and Strimmer, K. (2004). TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology*, **4**:18.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**:3059–3066.

Keays, M. C., Barker, D., Wicker-Thomas, C. and Ritchie, M. G. (2011). Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. *Molecular Ecology*, **20**:3617–3630.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**:111–120.

Li, W. H., Wu, C. I. and Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, **2**:150–174.

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, **290**:1151–1155.

Meesapyodsuk, D., Reed, D. W., Covello, P. S. and Qiu, X. (2007). Primary structure, regioselectivity, and evolution of the membrane-bound fatty acid desaturases of *Claviceps purpurea*. *The Journal of Biological Chemistry*, **282**:20191–20199.

Miyata, T. and Yasunaga, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its applications. *Journal of Molecular Evolution*, **16**:23–36.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**:418–426.

Nielsen, R. and Yang, Z. (1998). Likelihood models for detecting positively selected

amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**:929–936.

Nosil, P. and Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology and Evolution*, **26**:160–167.

Orr, H. A. (2005). The genetic basis of reproductive isolation: insights from *Drosophila*. *Proceedings of the National Academy of Sciences of the USA*, **102 Suppl 1**:6522–6526.

Perler, F., Efstratiadis, A. and Lomedica, P. (1980). The evolution of genes: the chicken preproinsulin gene. *Cell*, **20**:555–566.

Rouault, J., Capy, P. and Jallon, J.-M. (2000). Variations of male cuticular hydrocarbons with geoclimatic variables: an adaptive mechanism in *Drosophila melanogaster*? *Genetica*, **110**:117–130.

Schuerer, K. and Ledontal, C. (2003). protal2dna version 2.0. WWW. **URL:** *ftp://ftp.pasteur.fr/pub/gensoft/unix/alignment/protal2dna*

Sharp, P. M. and Li, W. H. (1989). An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution*, **24**:28–38.

Swanson, W. J., Nielsen, R. and Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, **20**:18–20.

Takahashi, A., Tsaur, S.-C., Coyne, J. A. and Wu, C.-I. (2001). The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **98**:3920–3925.

Tamura, K., Subramanian, S. and Kumar, S. (2004). Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Molecular Biology and Evolution*, **21**:36–44.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**:4673–4680.

Vieira, F. G., Sánchez-Garcia, A. and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology*, **8**:R235.

Wong, W. S. W., Yang, Z., Goldman, N. and Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, **168**:1041–1051.

Wu, C.-I. and Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, **5**:115–122.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, **13**:555–556.

Yang, Z. (1998). Likelihood ratio rests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, **15**:568–573.

Yang, Z. (2005). Paml faq. WWW.
URL: *http://abacus.gene.ucl.ac.uk/software/paml.html*

Yang, Z. (2006). *Computational Molecular Evolution*. Oxford: Oxford University Press.

Yang, Z. and Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution*, **46**:409–418.

Yang, Z. and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, **19**:908–917.

Yang, Z., Nielsen, R., Goldman, N. and Krabbe Pedersen, A.-M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**:431–449.

Yang, Z., Wong, W. S. W. and Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**:1107–1118.

Zhang, J., Nielsen, R. and Yang, Z. (2005). Evaluation of an improved branch-site method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**:2472–2479.

Chapter 4

Characterisation of a Novel Desaturase Gene

4.1 Introduction

In the work presented in Chapter 2, a new desaturase locus was detected in the two sequenced species of the *obscura* group, *D. pseudoobscura* and *D. persimilis*. It was not found in any other species. It is located on the X chromosome in *D. pseudoobscura*, and on chromosome 2 in *D. persimilis*. Its very high sequence identity with *desat1*, and the phylogenetic analysis in Chapter 2, indicate that it is a recent duplicate of this gene. Accordingly, it has been named *desat1b*. It was shown in Chapter 2 that the N-terminal region of the putative Desat1b protein is highly divergent from that of Desat1, the two amino acid sequences having only 31.4% identity in *D. pseudoobscura* (Figure 2.4).

In Chapter 3, it was found that the *desat1b* gene shows strong evidence of positive selection. The result of BEB analysis shows that most of the codons under positive selection are near to the $5'$ end of the gene – the N-terminus of the protein. This is the same region that is highly divergent from the ancestral *desat1* gene, and suggests that this divergence has been driven by adaptive evolution.

However, one line of evidence is not enough to prove this locus is actually undergoing adaptive evolution, or if the positively selected changes to the sequence are really important for the function of the enzyme. If the function of a gene of

interest, and the sites in question, is already known, then it is relatively easy to hypothesise about what the consequences of selection may be (Jensen *et al.*, 2007). In other studies of this kind, where a gene is found to be under positive selection using phylogenetic analyses, these tests have been followed up with laboratory work on the expression and function of the gene in question, establishing a link between the sites predicted to be under positive selection and changes in the structure and function of the protein (Jensen *et al.*, 2007; Yang, 2006). These studies have used *in vitro* methods such as site-directed mutagenesis to elucidate the functions of sites in question (Bielawski *et al.*, 2004; Ilvarsson *et al.*, 2003; Norrgård *et al.*, 2006; Sawyer *et al.*, 2005).

As yet, the function of Desat1b is not known. Uniprot shows that it has been automatically annotated in both species as being involved in fatty acid biosynthesis (GO:0006633), binding iron, being a transmembrane protein (GO:0016021) and having oxidoreductase activity (GO:0016491). It has been assigned the name GA23412 in *D. pseudoobscura* and GL15830 in *D. persimilis*. Given that it appears to be a duplicate of *desat1*, it is possible that it has a similar function, playing a role in the modification of cuticular hydrocarbons (CHCs). The gene does not contain any premature stop codons, and PAML has detected that the codon substitutions at the $5'$ end show a signal of positive selection, and not simply of relaxation of selective constraints. Therefore, it is likely that it is a functional desaturase gene. However, it is not known whether the *desat1b* locus is even transcribed, and so before it can be subjected to any functional experiments, the first step is to find out if it is. To do this, reverse transcriptase PCR was performed using RNA from male and female flies of *D. pseudoobscura* and *D. persimilis*. Following this, RACE PCR was used to confirm RT-PCR findings and determine the sequence of the $5'$ untranslated region (UTR). A summary of the work presented in this chapter appears in Keays *et al.* (2011).

4.2 Methods

4.2.1 RT-PCR

All flies were reared on standard cornmeal food medium (recipe in Appendix A, Section A.4) at 23°C with a 12:12 hour light/dark cycle. Total RNA was extracted from whole adult flies (20 males and 20 females each of *D. pseudoobscura* and *D. persimilis*) using the QIAGEN RNeasy® Mini kit (catalog number: 74106). For each RNA sample, twenty flies of the same sex were used. Live flies were placed in a 1.5 ml microcentrifuge tube and this was placed at −20°C for approximately five minutes to anaesthetise the flies. The tube was then opened and immersed slowly in liquid nitrogen, allowing some liquid to fall into the tube and cover the flies. A plastic pestle was also cooled in liquid nitrogen and then used to quickly crush the flies to powder. After the liquid nitrogen in the tube had evaporated, 600 µl Buffer RLT was immediately added to the tube, before the crushed tissue could thaw. From this point the protocol for Animal Tissues from the QIAGEN RNeasy Mini kit was followed, including the optional DNase incubation step between steps 6 and 7. Total RNA concentration was measured using a NanoDrop. The 50 µl sample was then split into five 10 µl aliquots and stored at −20°C.

cDNA synthesis

Complementary DNA (cDNA) was generated using the Bio-Rad iScript™ cDNA Synthesis kit (catalog number: 170-8891). The reaction was carried out in 0.2 ml PCR tubes. For each reaction, 250 ng total RNA was used. The volume containing this amount was calculated using the equation 4.1, where x is the volume to be calculated (in µl) and y is the concentration of the total RNA sample as measured by the NanoDrop (in ng/µl).

$$x = \frac{250}{y} \tag{4.1}$$

The cDNA synthesis reactions were set up as shown in Table 4.1. The "No RT" reaction was used as a control for contamination with genomic DNA. *Taq* DNA polymerase, to be used in the PCR step, is unable to amplify RNA under normal

PCR conditions (Myers and Gelfand, 1991). The "No RT" control reaction should not produce any cDNA, therefore no product should result. If product is seen, it must be because genomic DNA was not completely removed from the RNA sample, and brings into question the origin of any products generated from the cDNA sample. The "No RNA" control was used to guard against RNA or DNA contamination in the other reagents. Again, this sample should not give rise to any PCR products.

	Amount (µl)		
Reagent	cDNA synthesis	No RT	No RNA
5x iScript reaction mix	4	4	4
iScript reverse transcriptase	1	–	1
RNA template	1	1	–
Nuclease-free H_2O	14	15	15
Total	20	20	20

Table 4.1: Reagents for cDNA synthesis using Bio-Rad iScript kit.

The tubes were then incubated on a thermal cycler using the following program:

1. 5 minutes at 25°C
2. 30 minutes at 42°C
3. 5 minutes at 85°C

The reactions were then stored at −20°C.

PCR using cDNA

Two pairs of primers were designed, using Primer3 (version 0.4.0; Rozen and Skaletsky, 2000), to target the divergent region at the 5′ end of the gene (Figure 4.1). The forward primer of pair A is complementary to the sequence beginning 35 bp downstream of the putative start codon. The expected product size of this pair is 244 bp in both cDNA and genomic DNA. Primer pair B anneals further downstream in exon 1, with its forward primer target beginning 61 bp downstream of the putative start

codon. This primer in fact begins at a second in-frame ATG, which is another possible candidate for the start codon. The expected product size from primer pair B is 426 bp in both cDNA and genomic DNA. Primer sequences are given in Table 4.2. The rest of the *desat1b* gene, downstream of the 5′ divergent region, has extremely high sequence identity with its ancestor *desat1*. From the beginning of the putative start codon up to base pair 169, *desat1b* has 48.52% identity with *desat1*. Beyond this, from base pair 170 up to the putative stop codon, the identity increases to 80.64%. It was therefore impossible to design specific primers to target this region in *desat1b* only. Any primers binding to this region would also bind to *desat1*, so it would be impossible to determine which gene any resulting PCR products were from. As a positive control, primers were also designed to target *GapDH*.

1µl of each of the reactions from the cDNA synthesis was used as template in a PCR with each of the primer pairs (see Appendix A Section A.2 for protocol). Genomic DNA was also amplified to verify primer design, and 5µl each reaction was loaded onto a 2% agarose gel to visualise products.

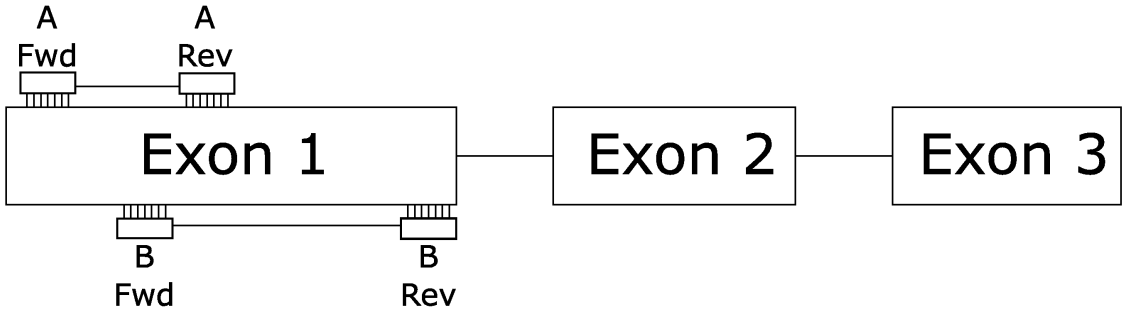


Figure 4.1: Primers used in RT-PCR of *desat1b*.

Pair	Forward primer	Reverse primer
A	AAGCAGTGGTAACCGTCGAT	TTGGCCGAAGTAACCAAAAG
B	ATGCAAGTGAATGCAGGAAC	CGTCTCCGAGAACTTGTGGT
GapDH	GCGCGGAATACGTAGTTGAA	GCGCACAGTTAAATCGACAA

Table 4.2: Sequences of primers used in RT-PCR. Annealing temperature for all primer pairs was 48.8°C

4.2.2 5′ RACE PCR

RNA was extracted from *D. pseudoobscura* and *D. persimilis* males and females using the QIAGEN RNeasy® Mini kit, as previously. The Clontech SMARTer™ RACE cDNA Amplification (Catalog No. 634923) and Advantage® 2 PCR (Catalog No. 639207) kits were used to generate RACE-ready cDNA and perform the RACE PCRs. The positive control steps using the Control Mouse Heart RNA provided in the cDNA Amplification kit were carried out first. RACE-ready cDNA and 5′ RACE PCR products were generated following the protocols in the SMARTer™ RACE cDNA Amplification Kit User Manual (Protocol No. PT4096-1). For the positive control RACE PCRs, all five reaction tubes in Table IV on page 17 of the User Manual were set up.

A gene-specific primer, GSP1, was designed for the 5′ RACE PCR from *D. pseudoobscura* and *D. persimilis* male and female cDNA. Prior to the RACE PCR, GSP1 was tested in PCRs on *D. pseudoobscura* and *D. persimilis* male and female genomic DNA, with another primer (GSP1_F) which targets the region 5′ of the putative start codon. The expected size of the product of this primer pair is 454 bp. Sequences of these primers are given in Table 4.3. Tubes 2 and 3 from Table IV on page 17 of the user manual were not set up for the RACE PCRs from *Drosophila* cDNA: Tube 2 is for mouse RNA only; Tube 3 is for overlapping 5′ and 3′ RACE products. 3′ RACE was not carried out. The RACE PCR was run using Program 1 on page 18 of the user manual.

Name	Sequence
GSP1	AGCAGTGATGCCAAGTCCAGAGCAC
GSP1_F	AAGCGAAGCTCAATCTGTCGTCTCG

Table 4.3: Sequences of primers for RACE PCR and testing in genomic DNA.

Following the RACE PCR reactions, all RACE products were gel-purified using the Clontech NucleoTrap® Gel Extraction Kit (included with RACE kit). The extracted RACE products were then used as template in a further round of PCR, using the protocol shown in Appendix A, Section A.2, GSP1 (10 µM) and the Universal

Primer Mix from the RACE kit, and an annealing temperature of 68°C. The products of these reactions were isolated using the Invitek Invisorb® Fragment CleanUp kit (Catalog No. 1020300300).

4.2.3 Cloning and sequencing of 5′ RACE products

RACE products were cloned using the Invitrogen™ TOPO® TA Cloning® Kit for Sequencing with One Shot® TOP10 Chemically Competent *E. coli* cells (Catalog No. K4575-01). Following the cloning kit protocol, the RACE products were ligated into the pCR®4-TOPO® vector (supplied with the cloning kit), which was then used to transform competent *E. coli* cells. Transformed cells were grown overnight at 37°C on LB/agar plates containing 50µg/ml ampicillin. The next day, ten colonies per plate were selected and cultured overnight at 37°C in liquid LB medium containing 50µg/ml ampicillin.

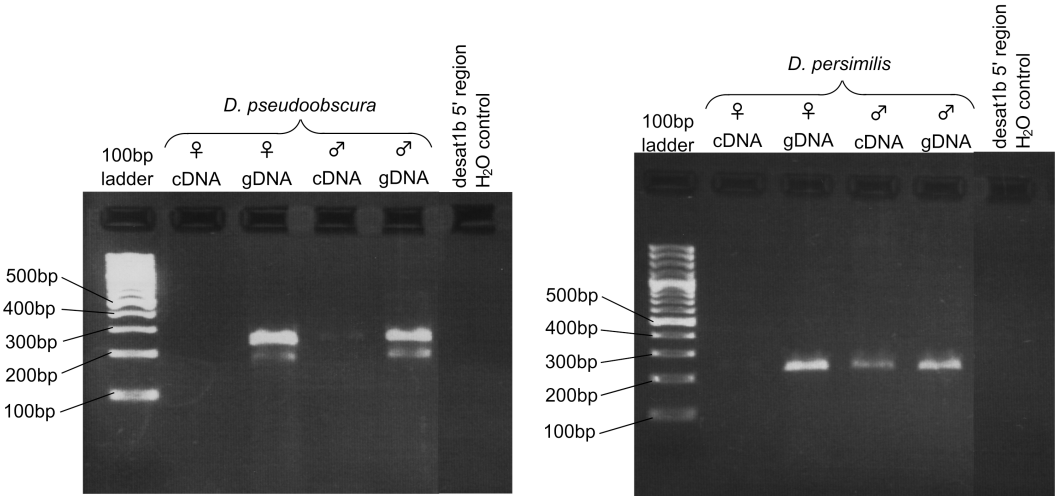
Plasmid DNA was extracted from the liquid cultures using the Invitrogen™ PureLink™ Quick Plasmid Miniprep Kit (Catalog No. K2100-10). The concentration of the extracted plasmid was ascertained using a NanoDrop. Plasmid DNA was sent for sequencing with M13 forward and reverse primers to the GenePool (Edinburgh, UK). Sequences were trimmed of pCR®4 vector, RACE kit universal primer and SMARTer™ II oligonucleotide sequences. They were then aligned using MAFFT (E-INS-i method; Katoh *et al.* (2002)) and manual adjustment.

4.3 Results

4.3.1 RT-PCR

The gels of the products from RT-PCR using primer pairs A and B (Table 4.2) are shown in Figures 4.2 and 4.3. All PCRs from genomic DNA produced strong bands of the expected size. In PCRs from cDNA, both primer pairs produced stronger bands from male cDNA than from female cDNA, for both *D. persimilis* and *D. pseudoobscura*, suggesting that the *desat1b* gene is more highly transcribed in males than in females. No amplification was detected in any of the No RT or No RNA

controls, indicating no contamination of genomic DNA in any of the samples.



(a) Products of primer pair A in *D. pseudoobscura*.

(b) Products of primer pair in *D. persimilis*.

Figure 4.2: Products of primer pair A in (a) *D. pseudoobscura* and (b) *D. persimilis*. The 5′ divergent region appears to be silent, or very weakly transcribed, in *D. pseudoobscura* females. It is weakly transcribed in *D. persimilis* females (very faint band). The region is transcribed much more strongly in males of both species. All amplified products are the expected size. Original images are of larger gels: only relevant lanes are shown here.

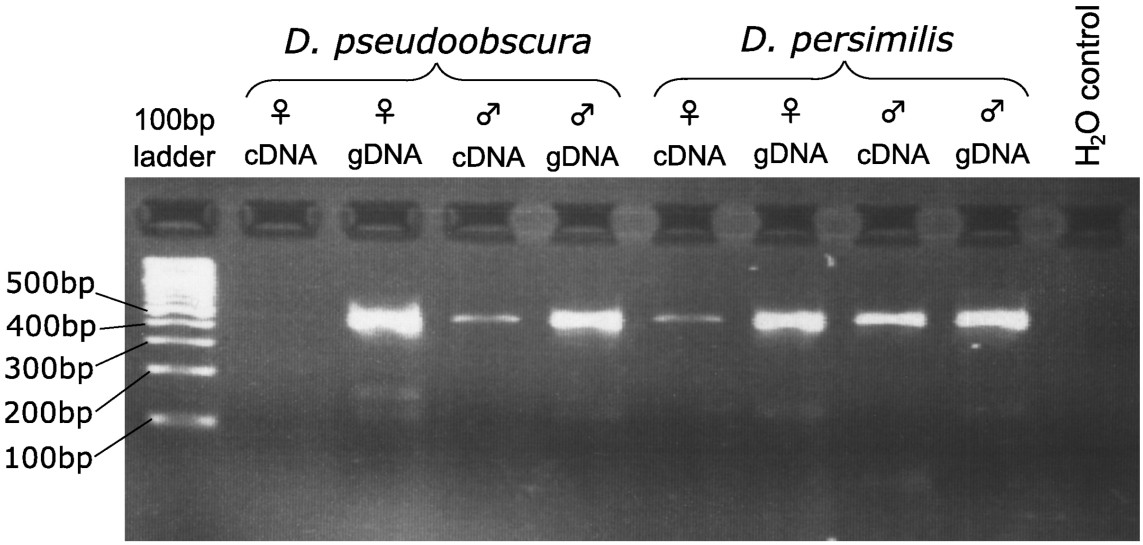


Figure 4.3: Products of primer pair B in *D. pseudoobscura* and *D. persimilis*. The product is extremely weak or non-existent in *D. pseudoobscura* female cDNA. As with primer pair A, it is stronger in male cDNA in both species. All amplified products are the expected size.

4.3.2 5′ RACE

Positive control

The positive control reactions using mouse heart RNA were successful, producing bands of 2.1 kb and 3.1 kb (Figure 4.4).

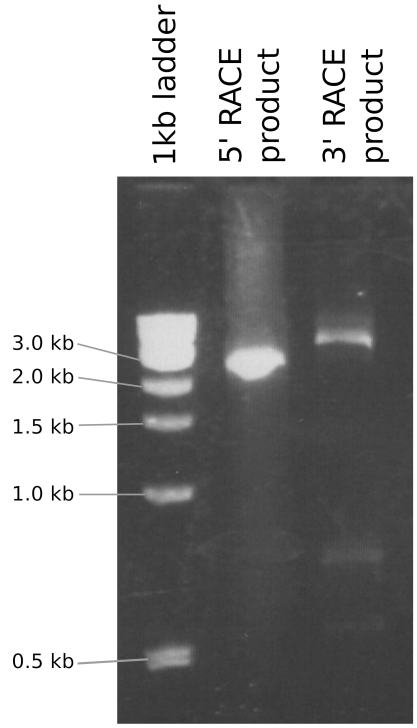


Figure 4.4: Gel of positive control RACE PCR using mouse heart RNA.

The PCRs using the GSP1 and GSP1_F primers in genomic DNA all produced strong bands of the expected size (454 bp), hence the GSP1 primer appears to be working and targeting the correct sequence (Figure 4.5).

The RACE PCRs for all samples were successful. The products of 5′ RACE in *D. pseudoobscura* male and female, and *D. persimilis* female samples are shown in Figures 4.6 and 4.7. The *D. persimilis* male product was isolated using the Invitek Invisorb® Fragment CleanUp kit only, as no non-specific bands were present (Figure 4.8). The extra round of PCR was not carried out for this sample, as the band was strong.

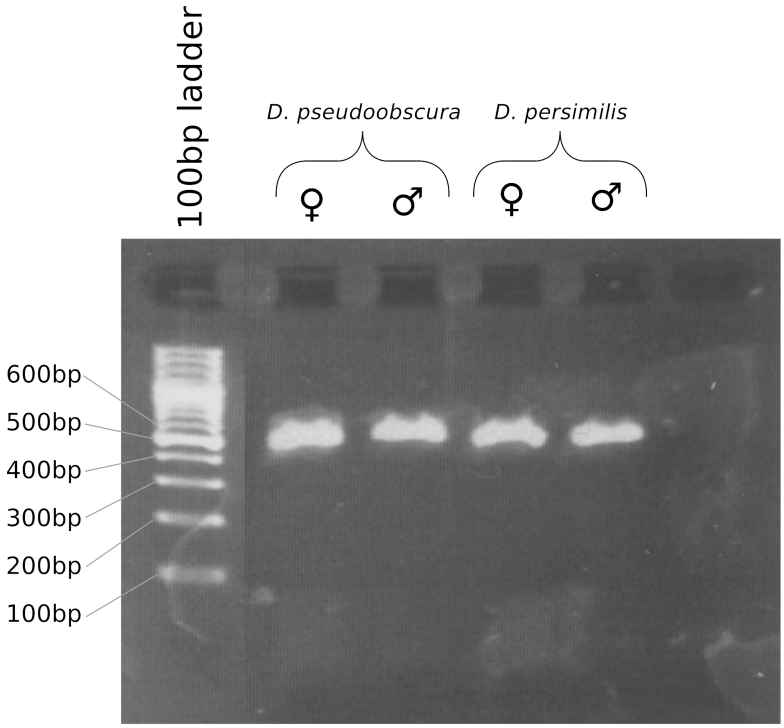


Figure 4.5: Gel of products of GSP1 and GSP1_F from genomic DNA. The PCRs all produced strong bands of the expected size.

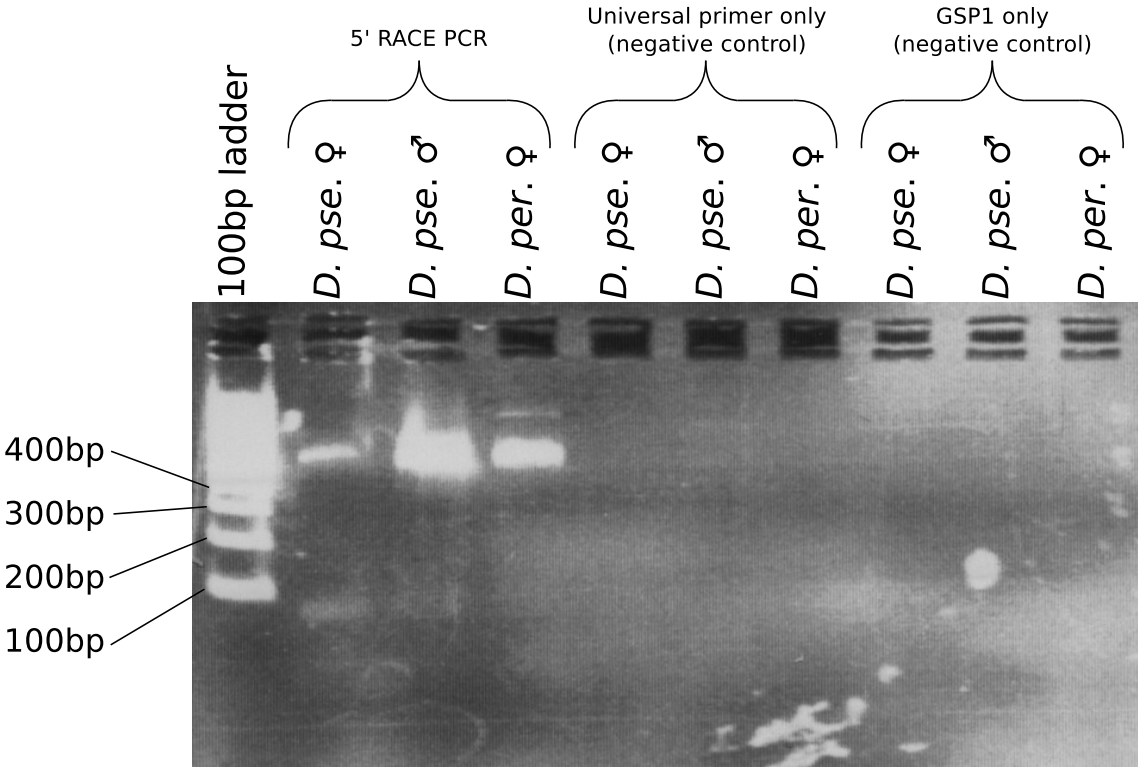
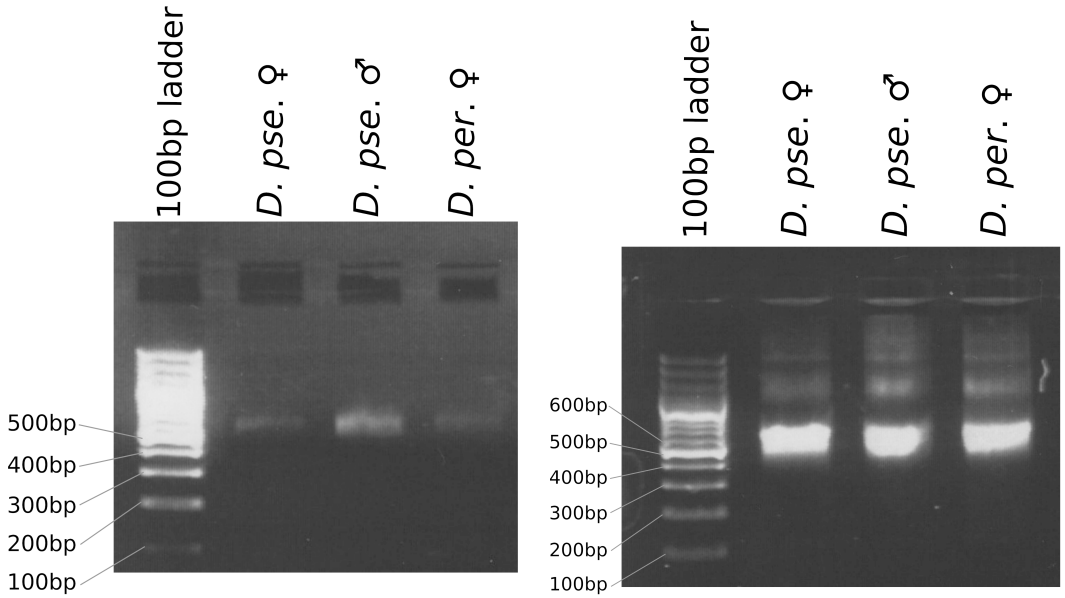


Figure 4.6: Gel of RACE products from *D. pseudoobscura* male and female, and *D. persimilis* female



(a) Gel of RACE products following gel extraction

(b) Gel of RACE products following second PCR

Figure 4.7: Gels of RACE products following gel-extraction and second round of PCR.

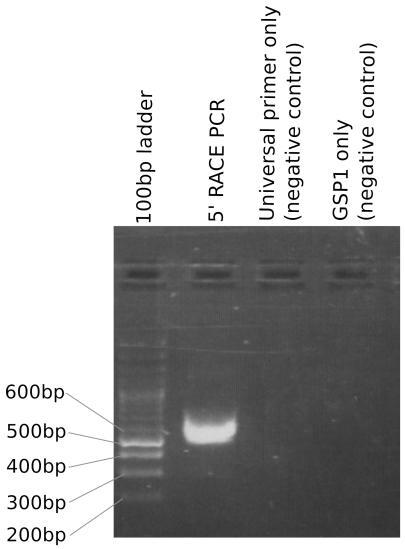


Figure 4.8: Gel of *D. persimilis* male RACE product.

4.3.3 Cloning and sequencing

It was not possible to extract high quality plasmid DNA from all ten colonies selected from each transformation plate: extractions from nine *D. pseudoobscura* female, four *D. pseudoobscura* male, ten *D. persimilis* female, and eight *D. persimilis* male cultures were successful. All successfully extracted plasmid samples were sent for sequencing with M13 forward and reverse primers. Of these, not all sequence traces were high quality: six forward and seven reverse primer sequences from *D. pseudoobscura* female, four forward and four reverse primer sequences from *D. pseudoobscura* male, ten forward and nine reverse primer sequences from *D. persimilis* female, and five forward and five reverse primer sequences from *D. persimilis* male samples were of sufficient quality to use in a multiple alignment. The multiple alignment of all high-quality RACE product sequences with the *D. pseudoobscura* and *D. persimilis desat1b* genomic sequence, including 1000 bp upstream of the putative start codon, is given in Appendix C.

The alignment shows that all of the RACE products contain the putative start codon (columns 1107–1109) and the divergent region immediately downstream of it, where positive selection was detected. The products also contain sequence 5′ of the putative start codon, aligning with the region immediately 5′ to the gene in the genomic DNA. In most of the products, this region extends for between 80 and 168 bp in the 5′ direction. In two of the *D. pseuodoobscura* male products (pse_male_1 and pse_male_4) however, the amount of sequence found that aligns in this region is shorter (19 bp). The rest of the putative 5′ UTR sequence aligns much further upstream in the genomic sequence. In both products this region begins 607 bp upstream of the start codon. In pse_male_1 it extends for 130 bp in the 5′ direction; in pse_male_4 it extends for 169 bp. This region has been designated the name "Exon 0". This suggests there is possible alternative splicing in the 5′ UTR of *desat1b* in *D. pseudoobscura* males.

One of the *D. pseudoobscura* male RACE products (pse_male_1) has "GTG" instead of "ATG" at the start codon position. Also, one of the *D. pseudoobscura* female RACE products appears to have a premature stop codon, "TGA", at columns 1387–1389. All other sequences have "TGG" at this position. The per_fem_10

forward and reverse sequences have a deletion of 45 bp, between columns 1259 and 1303 of the alignment. It is not clear whether these differences are real, or are a result of errors introduced at some stage during the PCR, cloning and sequencing processes. Three sequences, per_fem_5_fwd, pse_fem_02_fwd, and pse_fem_09_fwd, have relatively poor quality sequence traces. The "TGA" found in pse_fem_09_fwd is also present in pse_fem_09_rev, however, which has a high quality trace. So the "TGA" found in the pse_fem_09 sequences is not a result of low-quality sequence trace. The traces for per_fem_10 and pse_male_1 forward and reverse are all high-quality.

4.4 Discussion

The initial RT-PCR results show that the region in which positive selection was detected is actually transcribed. This strengthens the case for *desat1b* being a functional gene, and makes it more likely that the positive selection detected in Chapter 3 reflects strong selection such as may occur by neofunctionalisation following duplication. The RT-PCR results also suggest that the *desat1b* gene is under sex-specific differential expression regulation: it appears to be more strongly transcribed in males than it is in females (Figures 4.2 and 4.3).

Shirangi *et al.* (2009) found that the sex-specific expression of *desatF* is regulated by the Doublesex transcription factor, and that the gain and loss of *desatF* expression across the *Drosophila* is likely to be a direct result of the gain and loss of binding sites for this protein in the regulatory regions of *desatF*. Doublesex (DSX) is expressed in a sexually dimorphic manner in *Drosophila*, with male and female forms responsible for sex-biased expression of many genes (Burtis and Baker, 1989). Genes under DSX control have a cis-regulatory element (CRE) which contains the palindromic DSX target consensus sequence: $^{\mathrm{G}}_{\mathrm{A}}$NNAC$^{\mathrm{A}}_{\mathrm{T}}A^{\mathrm{T}}_{\mathrm{A}}$GTNN$^{\mathrm{C}}_{\mathrm{T}}$ (Erdman *et al.*, 1996). Along with other genes, such as *fruitless* and *transformer*, DSX is an important regulator of sex determination in *Drosophila*, responsible for sexually dimorphic behaviour and physiology (Siwicki and Kravitz, 2009; Verhulst *et al.*, 2010).

The 3000bp region upstream of the putative *desat1b* start codon in the genome

sequences of *D. pseudoobscura* and *D. persimilis* contains some sequences which are close matches to this consensus. The sequence GAGACAATGTATA was found in this region in both species. In *D. persimilis*, the sequence CTCACAATGTTTT was also detected. These sequences match the consensus of Erdman *et al.* (1996), minus the 5′- and 3′-most bases. It is as yet unknown whether this region forms part of a CRE, however. The sex-biased expression pattern of this gene suggests that the selection found in Chapter 3 is potentially the result of sexually divergent adaptive evolution.

It is difficult to tell from the RT-PCR gels alone whether the gene is at all transcribed in *D. pseudoobscura* females – the bands from cDNA are barely visible. The RACE PCRs have shown that the *desat1b* gene is transcribed in both sexes of both species, however, though before the second round of PCR the bands from females of both species were fainter than those from males. They also show that *desat1b* transcripts contain all sequence from the putative start codon downstream, including the divergent region, lending further support to the hypothesis that the positive selection is genuine. The RACE PCR products also contain some sequence upstream of the putative start codon – the putative 5′ UTR. The results of the RACE suggest that the 5′ UTR is alternatively spliced, at least in *D. pseudoobscura* males (see alignment in Appendix C). We cannot say if this also occurs in females of this species, or in either sex of *D. persimilis*. A next step in determining whether it does would be to design primers to target each of the UTRs found, and use these in RT-PCR on RNA from both sexes of both species. Also, Q-PCR could be employed to investigate the difference between male and female expression levels at a finer resolution, and to examine levels of transcripts possessing alternative 5′ UTRs.

Many genes have been found to have transcripts with alternatively spliced 5′ UTRs. In fact, *desat1b*'s closest relative, *desat1*, has five transcripts in *D. melanogaster*, differing only in their 5′ UTRs. Marcillac *et al.* (2005) found that four of the transcripts ("RA", "RB", "RC" and "RD") are expressed from larval stages through to adulthood, while the other one ("RE") was only present during metamorphosis. The functions of the different *desat1* transcripts are not yet known, but their tissue-specific expression is currently under investigation (Houot *et al.*, 2010). Hughes

(2006) reviews the various ways in which alternative UTRs have been found to be involved in regulation of translation in mammals. Alternative UTRs can determine tissue-specific expression patterns, such as in the gene *AXIN2*, which has three different 5′ UTRs. Each UTR provides the transcript with a different stability and translational efficiency. This means that the amount of axin2 protein produced in different tissues depends on the relative amounts of transcripts with each UTR: the higher the proportion of transcripts with UTRs enabling efficient translation, the more protein is produced, and vice versa (Hughes and Brady, 2005). The gene *FGF1* has four promoters giving rise to tissue-specific expression of four 5′ UTRs, which enable translation at different efficiencies and via different mechanisms in different cell types (Martineau *et al.*, 2004). Alternative 5′ UTRs have also been implicated in some cancers, for example in certain breast cancers, the expression of the gene *BRCA1* in downregulated as a result of a switch to a 5′ UTR that hampers translation (Sobczak and Krzyzosiak, 2002). It is therefore important to remember that the level of an mRNA transcript found does not necessarily reflect the level of protein that will be produced, as UTRs provide another level of gene regulation (Hughes, 2006). UTRs can also be targets of small non-coding regulatory RNAs, although this seems to affect 3′ UTRs more often than 5′ ones (Ulveling *et al.*, 2011).

Most of the sites under positive selection were shown to be located near the N-terminus of the Desat1b protein. It is not yet known whether *desat1b* mRNA is translated into a functional protein. The function of the N-terminus in *Drosophila* desaturase enzymes is as yet unclear. Desaturase enzymes do not act alone, rather they form an electron transport chain with NADH-cytochrome *b5* reductase and cytochrome *b5* (Nakamura and Nara, 2004). It is possible that the N-terminus is involved in interacting with these other proteins. Mziaut *et al.* (2000) found that the N-terminus of a mammalian desaturase, SCD, contains part of a rapid degradation signal. When part of the N-terminus was removed, the desaturase protein persisted for much longer than the fully intact enzyme. Concerning *desat1b*, the first step should be to determine whether the protein is actually produced at all. This could be done by creating an antibody to the Desat1b protein and performing a Western blot. A caveat for this is that Desat1b is very similar in sequence to Desat1, and

so care would need to be taken to select an antibody that is specific for Desat1b. If the Desat1b protein is found to be produced, the function of the N-terminus and the residues under selection can be investigated. Some studies have found that a particular locus has been subject to diversifying selection, and then have gone on to discover the functional effects of the changes this selection has caused (Yang, 2006). Ilvarsson *et al.* (2003) found evidence of positive selection in a human locus, glutathione transferase (GST). They then used site-directed mutagenesis *in vitro* to mutate the codons predicted to be under selection, and found that mutation of these sites greatly affected the specificity of the resulting enzyme. Bielawski *et al.* (2004) also used site-directed mutagenesis to investigate positively selected sites in a retinal-binding membrane protein in marine bacteria. This protein, proteorhodopsin, functions as a light-driven proton pump. The analysis performed by Bielawski *et al.* (2004) showed that the sites in question were important in light absorption sensitivity. For Desat1b this could be used to investigate effects on interactions with other proteins.

Another potential avenue of research into this gene could be to disrupt its function *in vivo* and examine the effects. Though *desat1b* is not found in the model species, *D. melanogaster*, ever larger numbers of studies are using non-model organisms to perform techniques such as gene knock-out or knock-down, and misexpression to infer the function of new genes (Shuker *et al.*, 2003; Terenius *et al.*, 2011). There is therefore a lot of scope for further work on the *desat1b* locus. It will be interesting to see whether, like *desat1*, it is also involved in pheromonal communication.

Bibliography

Bielawski, J. P., Dunn, K. A., Sabehi, G. and Beja, O. (2004). Darwinian adaptations of proteorhodopsin to different light intensities in the marine environment. *Proceedings of the National Academy of Sciences of the USA*, **101**:14824–14829.

Burtis, K. C. and Baker, B. S. (1989). *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, **56**:997–1010.

Erdman, S. E., Chen, H. J. and Burtis, K. C. (1996). Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila* doublesex proteins. *Genetics*, **144**:1639–1652.

Houot, B., Bousquet, F. and Ferveur, J. F. (2010). The consequences of regulation of desat1 expression for pheromone emission and detection in *Drosophila melanogaster*. *Genetics*, **185**:1297–1309.

Hughes, T. A. (2006). Regulation of gene expression by alternative untranslated regions. *Trends in Genetics*, **22**:119–122.

Hughes, T. A. and Brady, H. J. M. (2005). Expression of axin2 is regulated by the alternative 5′-untranslated regions of its mRNA. *Journal of Biological Chemistry*, **280**:8581–8588.

Ilvarsson, Y., Mackey, A. J., Edalat, M., Pearson, W. R. and Mannervik, B. (2003). Identification of residues in glutathione transferase capable of driving functional diversification in evolution. *Journal of Biological Chemistry*, **278**:8733–8738.

Jensen, J. D., Wong, A. and Aquadro, C. F. (2007). Approaches for identifying targets of positive selection. *Trends in Genetics*, **23**:568–577.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**:3059–3066.

Keays, M. C., Barker, D., Wicker-Thomas, C. and Ritchie, M. G. (2011). Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. *Molecular Ecology*, **20**:3617–3630.

Marcillac, F., Bousquet, F., Alabouvette, J., Savarit, F. and Ferveur, J. F. (2005). A mutation with major effects on *Drosophila melanogaster* sex pheromones. *Genetics*, **171**:1617–1628.

Martineau, Y., Bec, C. L. *et al.* (2004). Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNAs. *Molecular and Cellular Biology*, **24**:7622–7635.

Myers, T. W. and Gelfand, D. H. (1991). Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry*, **30**:7661–7666.

Mziaut, H., Korza, G. and Ozols, J. (2000). The N terminus of microsomal $\delta 9$ stearoyl-CoA desaturase contains the sequence determinant for its rapid degradation. *Proceedings of the National Academy of Sciences of the USA*, **97**:8883–8888.

Nakamura, M. T. and Nara, T. Y. (2004). Structure, function and dietary regulation of $\Delta 6$, $\Delta 5$, and $\Delta 9$ desaturases. *Annual Review of Nutrition*, **24**:345–376.

Norrgård, M. A., Ivarsson, Y., Tars, K. and Mannervik, B. (2006). Alternative mutations of a positively selected residue elicit gain or loss of functionalities in enzyme evolution. *Proceedings of the National Academy of Sciences of the USA*, **103**:4876–4881.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S. and Misener, S. (eds.), *Bioinformatics*

Methods and Protocols: Methods in Molecular Biology, pages 365–386. Totowa, New Jersey: Humana Press.

Sawyer, S. L., Wu, L. I., Emerman, M. and Malik, H. S. (2005). Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the USA*, **102**:2832–2837.

Shirangi, T. R., Dufour, H. D., Williams, T. M. and Carroll, S. B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biology*, **7**:e1000168.

Shuker, D., Lynch, J. and Morais, A. P. (2003). Moving from model to non-model organisms? lessons from *Nasonia* wasps. *BioEssays*, **25**:1247–1248.

Siwicki, K. K. and Kravitz, E. A. (2009). Fruitless, doublesex and the genetics of social behavior in *Drosophila melanogaster*. *Current Opinion in Neurobiology*, **19**:200–206.

Sobczak, K. and Krzyzosiak, W. J. (2002). Structural determinants of BRCA1 translational regulation. *Journal of Biological Chemistry*, **277**:17349–17358.

Terenius, O., Papanicolaou, A. *et al.* (2011). RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. *Journal of Insect Physiology*, **57**:231–245.

Ulveling, D., Francastel, C. and Hubé, F. (2011). When on is better than two: RNA with dual functions. *Biochimie*, **93**:633–644.

Verhulst, E. C., van de Zande, L. and Beukeboom, L. W. (2010). Insect sex determination: it all evolves around *transformer*. *Current Opinion in Genetics and Development*, **20**:376–383.

Yang, Z. (2006). *Computational Molecular Evolution*. Oxford: Oxford University Press.

Chapter 5

RNA Interference of Desaturase Gene Expression

5.1 Introduction

Several genes identified in Chapter 2 as being members of the desaturase family are as yet functionally uncharacterised. None of them have yet been implicated in the production of CHCs by methods used so far, such as QTL analysis. If they are functional desaturases, we might predict them to have a role in CHC modification. Various methods exist for manipulating gene expression *in vivo*, suppressing it so that the gene's function is removed, amplifying it so that the gene is overexpressed, or even causing ectopic expression of the gene (Roman, 2004). By observing the effects of these manipulations on living organisms, researchers can make inferences about the function of the gene. RNA interference (RNAi) is one method of suppressing a gene's expression.

Silencing of gene expression by interfering RNAs was first discovered in plants (Napoli *et al.*, 1990; van der Krol *et al.*, 1990; Voinnet, 2001). Fire *et al.* (1998) discovered that a similar phenomenon is present in the nematode worm *Caenorhabditis elegans*. They found that injecting double-stranded RNA (dsRNA) into adults of *C. elegans* resulted in suppression of gene expression. Genes with sequence matching the dsRNA were being silenced. This phenomenon is in fact conserved between many organisms, including insects, mammals and fungi (Hannon, 2002). The

RNAi system provides adaptive, sequence-specific defence against infection by RNA viruses. This has been demonstrated in plants (Vance and Vaucheret, 2001; Voinnet, 2001), and more recently in *C. elegans* (Wilkins *et al.*, 2005). The presence of dsRNA in a cell attracts the Dicer enzyme complex, which cuts it into short, double-stranded fragments. In *Drosophila*, these fragments are then bound by another enzyme complex known as RISC (RNA-induced silencing complex). When activated by ATP, RISC unwinds the dsRNA fragment and this then enables it to target single-stranded RNA species with complementary sequence via Watson-Crick base pairing. Targeted RNAs are degraded by RISC, thus preventing their translation. In this way, the RNAi pathway implements sequence-specific suppression of gene expression (Hannon, 2002).

The RNAi system can be invoked experimentally to silence expression of a gene of interest by injection, as mentioned above, however a drawback of injection is that the effects caused are transient, and not stably inherited by future generations. Kennerdell and Carthew (2000) found a way to overcome this using an inverted repeat (IR) sequence, which produced RNA with a hairpin-loop structure. They first tested the effiency of the hairpin-loop at inducing RNAi using *Drosophila* embryos expressing the *lacZ* gene under the control of the *engrailed* promoter, and injecting either a hairpin-loop or linear dsRNA, containing sequence corresponding to a portion of the *lacZ* gene. They found that the hairpin was as efficient at suppressing *lacZ* function as the dsRNA. They then created a transgene which contained the hairpin sequence, plus a yeast upstream activating sequence (UAS). UAS is the target of the yeast transcription factor GAL4. They generated *Drosophila* lines which contained these *UAS-IRlacZ* transgenes, *GAL4* under the control of the heat-shock-inducible *hsp70* promoter, and also a *UAS-lacZ* construct. After heat-shocking, flies that carried all three constructs showed greatly reduced *lacZ* expression compared with those that did not possess the hairpin-producing transgene, *UAS-IRlacZ*, indicating that the hairpin produced by the transgene was capable of invoking the RNAi mechanism. The main steps involved in GAL4-mediated RNAi are shown in Figure 5.1.

It is now possible to purchase *D. melanogaster* lines carrying a *UAS-IR* hairpin-producing construct corresponding to virtually any gene of interest, from organi-

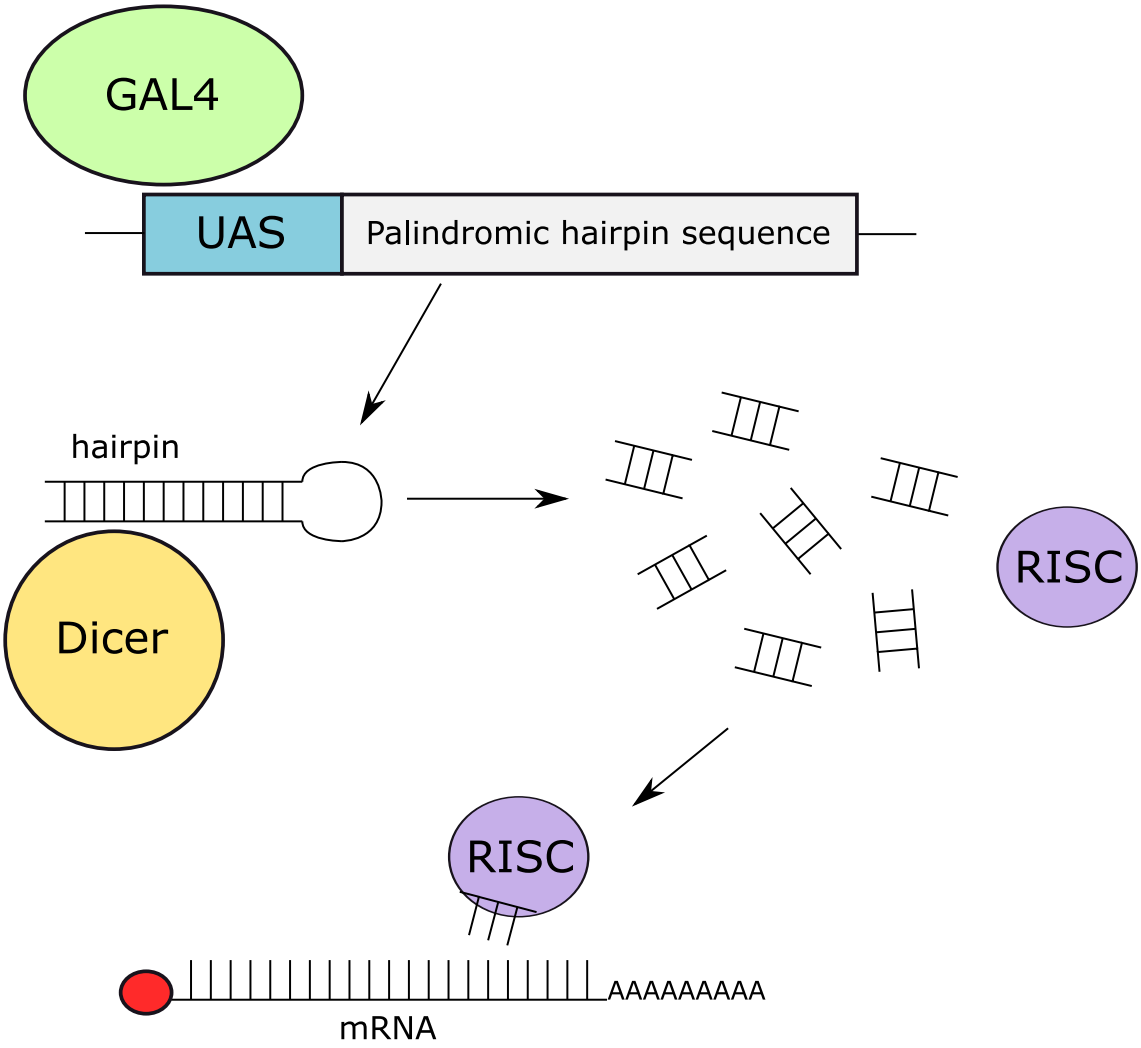


Figure 5.1: The main steps in GAL4-mediated RNAi. GAL4 binds to the UAS and initiates transcription of the palidromic sequence, which produces mRNA with a hairpin-loop structure. The hairpin is targeted by Dicer, which breaks it into double-stranded fragments 19-22 nucleotides long. These fragments are bound by RISC, which unwinds them to create single-stranded fragments. The RISC-bound single-stranded fragment allows specific targeting of mRNA (Hannon, 2002).

sations such as the Vienna *Drosophila* RNAi Center (`http://stockcenter.vdrc.at`) and *Drosophila* Genetic Resource Center in Kyoto (`http://kyotofly.kit.jp/cgi-bin/stocks/data_search.cgi`). Lines carrying GAL4 under various different promoters are also readily available from multiple sources, and as such it is easy to induce RNAi "knock-down" of any *D. melanogaster* gene in any specific tissue. This is done by simply crossing a line expressing the GAL4 driver in the desired tissue(s), with a line carrying the *UAS-IR* construct targeting the gene of interest. The F1 hybrids will carry both the *GAL4* gene and the *UAS-IR* construct, and thus the gene will be knocked-down in these individuals, only in the tissue(s) desired (Duffy, 2002).

RNAi has been used in numerous functional genetics studies in many different organisms (reviewed by Perrimon *et al.*, 2010). Maeda *et al.* (2001) used high-throughput RNAi to perform genome-wide functional analysis in *C. elegans*, and identified 24 genes with important roles in developmental pathways. More recently in the same organism Sönnichsen *et al.* (2005) performed another genome-wide screen, using RNAi to identify all the genes involved in the first two rounds of cell division. RNAi knock-down has also been employed in many studies of Lepidoptera, reviewed by Terenius *et al.* (2011), to advance understanding in processes including development and immunity. Genome-wide studies have been carried out in *Drosophila*: Mummery-Widmer *et al.* (2009) performed a genome-wide screen using transgenic RNAi to successfully identify regulators of the Notch signalling pathway; and Cronin *et al.* (2009) discovered numerous genes important for innate immumity and defence against bacterial infection in *Drosophila*, again using a genome-wide RNAi screen.

RNAi has also been used to investigate *Drosophila* desaturase gene function. Wicker-Thomas *et al.* (2009) silenced expression of *desat1* and *desatF* in the oenocytes, the cells in which CHC synthesis takes place. They used the 1407-GAL4 line, which expresses GAL4 in the oenocytes only. They found highly significant changes in CHC profiles in the resulting progeny, with a 96% and 78% decrease in unsaturated CHCs in males and females, respectively, in the *desat1* experiment, and a 98% decrease in dienes in females in the *desatF* analysis. In the work presented in this chapter, the same *GAL4* driver was used to induce RNAi knock-down of the seven

remaining desaturase genes identified in Chapter 2 – *desat2*, *CG8630*, *CG15531*, *CG17928*, *Cyt-b5-r*, *CG9743* and *CG9747*. The CHC profiles of the F1 progeny were examined to determine whether the knock-down has had any effect. If a desaturase is involved in CHC modification and its function is turned off, the expectation is that there would be a fall in the amount of unsaturated CHCs, as demonstrated by Wicker-Thomas *et al.* (2009). The ratios of unsaturated to saturated CHCs were therefore examined here, as well as the data for each individual CHC.

5.2 Methods

5.2.1 Crosses

All *D. melanogaster* lines used were kindly supplied by C. Wicker-Thomas. The GAL4 line used was *1407/Cy*, which is oenocyte-specific, and was created by C. Wicker-Thomas by crossing the *1407-GAL4* line with a *w;Cy;TM3* line. Only flies with curly wings are kept each generation. UAS lines used were *UAS-desat2-i*, *UAS-CG8630-i*, *UAS-CG15531-i*, *UAS-CG17928-i*, *UAS-Cyt-b5-r-i*, *UAS-CG9743-i* and *UAS-CG9747-i*. Lines were originally obtained from the Vienna *Drosophila* RNAi Center (Dietzl *et al.*, 2007). All lines apart from *UAS-CG15531-i* were homozygous-viable with the *UAS-IR* transgene inserted on chromosome II; *UAS-CG15531-i* was homozygous lethal with the transgene on chromosome II, and kept over the *CyO* balancer. Flies were reared on standard cornmeal food medium (recipe in Appendix A, Section A.4) at 23°C with a 12:12 hour light/dark cycle.

Figure 5.2 illustrates the crosses performed. For each *GAL4/UAS-IR* cross, eight virgin females with curly wings were selected from the *1407/CyO* line. These were added to a vial along with approximately eight virgin males of one *UAS-IR* line. All adults were removed after seven days. F1 adults began to appear on day 11; these were removed twice daily at 9:00 and 18:00, and males and females stored in separate vials. The F1 generation has two types of flies: those with straight wings and those with curly. The straight-winged flies have the genotype *UAS-IR/1407*, meaning they have both the *GAL4* driver and the *UAS-IR* construct. The RNAi system should therefore be active in oenocytes of these flies. The curly-winged flies

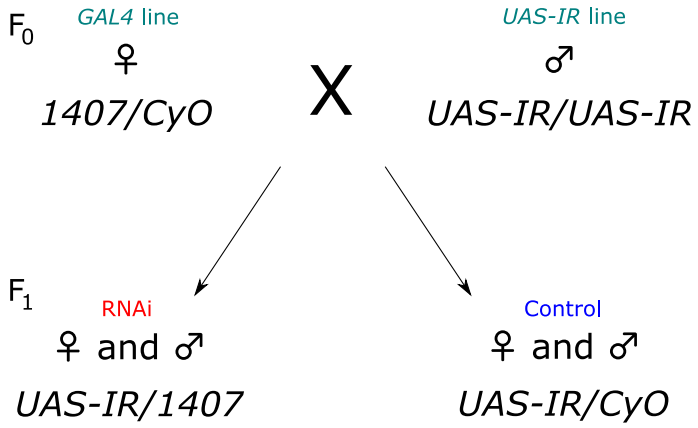


Figure 5.2: Diagram to show cross between a *GAL4* line and a *UAS-IR* line. F_1 hybrids with the genotype *UAS-IR/1407* will have activated RNAi in the oenocytes. Those with *UAS-IR/CyO* do not have *GAL4* and serve as controls.

do not have the *GAL4* driver and their genotype is *UAS-IR/CyO*. They serve as the negative control for the experiment.

5.2.2 CHC extractions

CHCs were extracted from flies at 24 hours old, and four days old, as the CHC profiles of *Drosophila* are known to change over time (Antony and Jallon, 1982; Ferveur, 2005; Pechine *et al.*, 1988). Forty flies were analysed from each cross: five males and five females of *UAS/1407* and five males and five females of *UAS/Cy* at 24 hours old; and the same numbers of males and females of each genotype at four days old. Live flies were removed from vials under CO_2 anaesthesia and placed at -20°C in an 1.5ml microcentrifuge tube for at least 10 minutes before beginning the extraction procedure.

The extraction procedure was carried out in a fume cupboard at room temperature. CHCs were extracted by submerging each single fly in 150µl 5µg/ml C26/heptane solution, in a cylindrical glass tube. Each fly was left in the solution for 10 minutes and then removed to a 1.5ml microcentrifuge tube containing 100% ethanol. Extracted flies in 100% ethanol were stored at -20°C. The glass tubes containing the solution were placed in glass bottles and left open to allow all liquid to evaporate. They were then sealed using plastic lids with a Teflon inlay and

sent to C. Wicker-Thomas who performed gas chromatography-mass spectrometry (GC-MS) analysis. The extract is injected into a gas chromatography (GC) column, which separates out the individual CHCs, based on size. Each CHC is then identified by mass spectrometry (MS), which breaks up the molecules and records the mass to charge ratio for each fragment. The GC-MS instrument produces a chromatogram with peaks corresponding to each individual CHC. The area of each peak corresponds to the amount of CHC present in the mixture. The absolute amount is calculated using the peak area of a standard, for which the amount present is known, as a reference (Kitson *et al.*, 1996).

5.2.3 Statistics

The GC-MS analysis produced spreadsheets containing the peak areas for each hydrocarbon, as well as the amount for each CHC in nanograms, estimated using the peak area and known amount of a C26 standard (equation 5.1).

$$\text{CHC peak area} \times \frac{\text{ng standard}}{\text{standard peak area}} \tag{5.1}$$

For each cross, the amounts of each CHC in ng were used to perform principal components analysis (PCA) in Minitab® (Minitab Inc.). CHCs absent from all individuals were removed from the data prior to analysis. Data were standardised to have means of zero and standard deviations of 1. Linear models were fitted to each principal component which explained greater than 5% of the total variance and/or had eigenvalues greater than 1. In the model, the principal component was used as the response variable and explanatory variables were Treatment (RNAi or Control; categorical), Age (1-day-old or 4-days-old; categorical) and Sex (Male or Female; categorical). The interactions between Treatment and Age and Treatment and Sex were also included in the model.

Because desaturases are responsible for levels of unsaturation in CHCs, in addition analysis of the ratio of unsaturated to saturated CHCs was performed, again using a linear model. The amounts of all dienes and all monoenes were combined to give the total amount of unsaturated CHCs for each individual. This was divided by the total amount of linear CHCs for each fly, to give the ratio of unsatu-

rated:saturated CHCs (equation 5.2). The ratio was used as the response variable, with the same explanatory variables as in the analysis of principal components. To correct for multiple testing, a P-value cut-off of 0.007 was used to detect statistical significance (equation 5.3).

$$\text{unsaturated:saturated CHC ratio} = \frac{\text{total monoenes} + \text{total dienes}}{\text{total linear CHCs}} \tag{5.2}$$

$$\frac{0.05}{7 \text{ genes}} = 0.007 \tag{5.3}$$

5.3 Results

5.3.1 Principal components analysis

ANOVA results for all PCAs are shown in Tables 5.1 to 5.40. Significant P values are in bold in the tables. Several P-values were less than 0.05; these "almost significant" values are shown in italics in the tables. After correcting for multiple testing, only three P values remained significant: Treatment (RNAi or Control) was found to have a significant effect on PC5 (4.4% variance) of the *CG8630* experiment; the Treatment*Age interaction also was also significant in PC5 for this gene. The Treatment*Sex interaction was significant in PC4 (6.0% variance) of the *CG15531* analysis.

desat2

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	13.242	5.672	1.999	1.701	1.120
% variance	47.3	20.3	7.1	6.1	4.0

Table 5.1: Eigenvalues and % variance explained for principal components in *desat2* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.124	0.124	0.03	0.866
Sex	1	265.206	265.206	61.79	**<0.001**
Age	1	103.490	103.490	24.11	**<0.001**
Treatment*Sex	1	0.501	0.501	0.12	0.735
Treatment*Age	1	1.192	1.192	0.28	0.602
Error	34	145.929	4.292		
Total	39	516.441			

Table 5.2: Analysis of variance for principal component 1 in *desat2* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.364	0.364	0.31	0.582
Sex	1	40.108	40.108	34.03	**<0.001**
Age	1	140.445	140.445	119.15	**<0.001**
Treatment*Sex	1	0.010	0.010	0.01	0.926
Treatment*Age	1	0.217	0.217	0.18	0.671
Error	34	40.077	1.179		
Total	39	221.221			

Table 5.3: Analysis of variance for principal component 2 in *desat2* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	9.403	9.403	5.36	0.027
Sex	1	0.176	0.176	0.10	0.753
Age	1	0.757	0.757	0.43	0.516
Treatment*Sex	1	5.004	5.004	2.85	0.100
Treatment*Age	1	2.987	2.987	1.70	0.201
Error	34	59.642	1.754		
Total	39	77.969			

Table 5.4: Analysis of variance for principal component 3 in *desat2* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	1.266	1.266	0.90	0.350
Sex	1	8.034	8.034	5.70	*0.023*
Age	1	0.411	0.411	0.29	0.593
Treatment*Sex	1	0.420	0.420	0.30	0.589
Treatment*Age	1	8.261	8.261	5.86	*0.021*
Error	34	47.955	1.410		
Total	39	66.347			

Table 5.5: Analysis of variance for principal component 4 in *desat2* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	7.2198	7.2198	7.54	*0.010*
Sex	1	1.2599	1.2599	1.32	0.259
Age	1	0.0350	0.0350	0.04	0.850
Treatment*Sex	1	1.1809	1.1809	1.23	0.275
Treatment*Age	1	1.4243	1.4243	1.49	0.231
Error	34	32.5724	0.9580		
Total	39	43.6923			

Table 5.6: Analysis of variance for principal component 5 in *desat2* analysis

CG8630

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	12.391	6.595	2.292	1.585	1.197
% variance	45.9	24.4	8.5	5.9	4.4

Table 5.7: Eigenvalues and % variance explained for principal components in *CG8630* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.121	0.121	0.03	0.864
Sex	1	271.316	271.316	67.24	**<0.001**
Age	1	73.742	73.742	18.27	**<0.001**
Treatment*Sex	1	0.620	0.620	0.15	0.697
Treatment*Age	1	0.265	0.265	0.07	0.799
Error	34	137.195	4.035		
Total	39	483.259			

Table 5.8: Analysis of variance for principal component 1 in *CG8630* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	5.441	5.441	3.30	0.078
Sex	1	37.425	37.425	22.69	**<0.001**
Age	1	151.009	151.009	91.54	**<0.001**
Treatment*Sex	1	0.755	0.755	0.46	0.503
Treatment*Age	1	6.475	6.475	3.93	0.056
Error	34	56.085	1.650		
Total	39	257.191			

Table 5.9: Analysis of variance for principal component 2 in *CG8630* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.130	0.130	0.06	0.815
Sex	1	0.254	0.254	0.11	0.744
Age	1	7.202	7.202	3.08	0.088
Treatment*Sex	1	1.964	1.964	0.84	0.366
Treatment*Age	1	0.297	0.297	0.13	0.724
Error	34	79.539	2.339		
Total	39	89.386			

Table 5.10: Analysis of variance for principal component 3 in *CG8630* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	9.439	9.439	0.72	0.009
Sex	1	0.815	0.815	0.66	0.421
Age	1	1.174	1.174	0.96	0.335
Treatment*Sex	1	8.144	8.144	6.64	*0.014*
Treatment*Age	1	0.554	0.554	0.45	0.506
Error	34	41.690	1.226		
Total	39	61.815			

Table 5.11: Analysis of variance for principal component 4 in *CG8630* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	8.2669	8.2669	11.73	**0.002**
Sex	1	3.0717	3.0717	4.36	*0.044*
Age	1	1.0612	1.0612	1.51	0.228
Treatment*Sex	1	0.0044	0.0044	0.01	0.938
Treatment*Age	1	10.3260	10.3260	14.65	**0.001**
Error	34	23.9712	0.7050		
Total	39	46.7012			

Table 5.12: Analysis of variance for principal component 5 in *CG8630* analysis

CG9743

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	12.876	6.551	1.601	1.322	1.112
% variance	47.7	24.3	5.9	4.9	4.1

Table 5.13: Eigenvalues and % variance explained for principal components in *CG9743* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.017	0.017	0.00	0.953
Sex	1	285.539	285.539	58.49	**<0.001**
Age	1	50.374	50.374	10.32	**0.003**
Treatment*Sex	1	0.043	0.043	0.01	0.926
Treatment*Age	1	0.218	0.218	0.04	0.834
Error	34	165.989	4.882		
Total	39	502.198			

Table 5.14: Analysis of variance for principal component 1 in *CG9743* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.045	0.045	0.04	0.852
Sex	1	47.750	47.750	37.50	**<0.001**
Age	1	164.055	164.055	128.83	**<0.001**
Treatment*Sex	1	0.157	0.157	0.12	0.728
Treatment*Age	1	0.173	0.173	0.14	0.715
Error	34	43.295	1.273		
Total	39	255.475			

Table 5.15: Analysis of variance for principal component 2 in *CG9743* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	8.476	8.476	5.97	0.020
Sex	1	4.141	4.141	2.92	0.097
Age	1	0.215	0.215	0.15	0.700
Treatment*Sex	1	0.772	0.772	0.54	0.466
Treatment*Age	1	0.572	0.572	0.40	0.530
Error	34	48.282	1.420		
Total	39	62.456			

Table 5.16: Analysis of variance for principal component 3 in *CG9743* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	1.143	1.143	0.78	0.384
Sex	1	0.222	0.222	0.15	0.700
Age	1	0.006	0.006	0.00	0.948
Treatment*Sex	1	0.005	0.005	0.00	0.955
Treatment*Age	1	0.222	0.222	0.15	0.700
Error	34	49.965	1.470		
Total	39	51.564			

Table 5.17: Analysis of variance for principal component 4 in *CG9743* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.829	0.829	0.77	0.385
Sex	1	0.028	0.028	0.03	0.874
Age	1	0.852	0.852	0.80	0.379
Treatment*Sex	1	4.576	4.576	4.27	*0.046*
Treatment*Age	1	0.631	0.631	0.59	0.448
Error	34	36.440	1.072		
Total	39	43.356			

Table 5.18: Analysis of variance for principal component 5 in *CG9743* analysis

CG9747

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	10.202	6.443	2.002	1.598	1.292
% variance	37.8	23.9	7.4	5.9	4.8

Table 5.19: Eigenvalues and % variance explained for principal components in *CG9747* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.888	0.888	0.28	0.603
Sex	1	265.265	265.265	82.32	**<0.001**
Age	1	21.306	21.306	6.61	*0.015*
Treatment*Sex	1	0.154	0.154	0.05	0.828
Treatment*Age	1	0.682	0.682	0.21	0.648
Error	34	109.564	3.222		
Total	39	397.860			

Table 5.20: Analysis of variance for principal component 1 in *CG9747* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	2.019	2.019	1.28	0.265
Sex	1	12.042	12.042	7.66	**0.009**
Age	1	183.084	183.084	116.39	**<0.001**
Treatment*Sex	1	0.052	0.052	0.03	0.857
Treatment*Age	1	0.607	0.607	0.39	0.539
Error	34	53.481	1.573		
Total	39	251.285			

Table 5.21: Analysis of variance for principal component 2 in *CG9747* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.028	0.028	0.01	0.907
Sex	1	2.858	2.858	1.40	0.245
Age	1	0.000	0.000	0.00	0.999
Treatment*Sex	1	5.727	5.727	2.81	0.103
Treatment*Age	1	0.084	0.084	0.04	0.841
Error	34	69.384	2.041		
Total	39	78.080			

Table 5.22: Analysis of variance for principal component 3 in *CG9747* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.399	0.399	0.24	0.625
Sex	1	0.005	0.005	0.00	0.956
Age	1	0.067	0.067	0.04	0.841
Treatment*Sex	1	1.803	1.803	1.10	0.302
Treatment*Age	1	4.244	4.244	2.59	0.117
Error	34	55.811	1.641		
Total	39	62.330			

Table 5.23: Analysis of variance for principal component 4 in *CG9747* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	5.197	5.197	5.11	0.030
Sex	1	0.000	0.000	0.00	0.984
Age	1	1.352	1.352	1.33	0.257
Treatment*Sex	1	3.558	3.558	3.50	0.070
Treatment*Age	1	5.900	5.900	5.80	0.022
Error	34	34.592	1.017		
Total	39	50.599			

Table 5.24: Analysis of variance for principal component 5 in *CG9747* analysis

CG15531

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	12.931	5.525	2.383	1.681	1.172
% variance	46.2	19.7	8.5	6.0	4.2

Table 5.25: Eigenvalues and % variance explained for principal components in *CG15531* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	2.854	2.854	0.66	0.423
Sex	1	205.363	205.363	47.39	**<0.001**
Age	1	146.115	146.115	33.72	**<0.001**
Treatment*Sex	1	1.713	1.713	0.40	0.534
Treatment*Age	1	0.935	0.935	0.22	0.645
Error	34	147.332	4.333		
Total	39	504.312			

Table 5.26: Analysis of variance for principal component 1 in *CG15531* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.627	0.627	0.45	0.508
Sex	1	39.306	39.306	28.09	**<0.001**
Age	1	126.279	126.279	90.24	**<0.001**
Treatment*Sex	1	0.054	0.054	0.04	0.845
Treatment*Age	1	1.621	1.621	1.16	0.289
Error	34	47.580	1.399		
Total	39	215.467			

Table 5.27: Analysis of variance for principal component 2 in *CG15531* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	9.116	9.116	4.07	0.052
Sex	1	2.607	2.607	1.16	0.288
Age	1	1.515	1.515	0.68	0.416
Treatment*Sex	1	1.726	1.726	0.77	0.386
Treatment*Age	1	1.858	1.858	0.83	0.369
Error	34	76.095	2.238		
Total	39	92.918			

Table 5.28: Analysis of variance for principal component 3 in *CG15531* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	10.469	10.469	10.38	**0.003**
Sex	1	1.503	1.503	1.49	0.231
Age	1	0.036	0.036	0.04	0.852
Treatment*Sex	1	17.603	17.603	17.45	**<0.001**
Treatment*Age	1	1.650	1.650	1.64	0.210
Error	34	34.296	1.009		
Total	39	65.556			

Table 5.29: Analysis of variance for principal component 4 in *CG15531* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.063	0.063	0.06	0.810
Sex	1	2.663	2.663	2.47	0.125
Age	1	0.007	0.007	0.01	0.938
Treatment*Sex	1	0.586	0.586	0.54	0.466
Treatment*Age	1	5.773	5.773	5.36	0.027
Error	34	36.610	1.077		
Total	39	45.701			

Table 5.30: Analysis of variance for principal component 5 in *CG15531* analysis

CG17928

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	11.018	7.006	2.836	1.672	1.102
% variance	40.8	25.9	10.5	6.2	4.1

Table 5.31: Eigenvalues and % variance explained for principal components in *CG17928* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	3.314	3.314	0.81	0.376
Sex	1	187.517	187.517	45.64	**<0.001**
Age	1	92.763	92.763	22.58	**<0.001**
Treatment*Sex	1	4.411	4.411	1.07	0.307
Treatment*Age	1	1.983	1.983	0.48	0.492
Error	34	139.700	4.109		
Total	39	429.689			

Table 5.32: Analysis of variance for principal component 1 in *CG17928* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	1.837	1.837	1.45	0.236
Sex	1	51.796	51.796	41.01	**<0.001**
Age	1	171.264	171.264	135.59	**<0.001**
Treatment*Sex	1	4.707	4.707	3.73	0.062
Treatment*Age	1	0.680	0.680	0.54	0.468
Error	34	42.945	1.263		
Total	39	273.229			

Table 5.33: Analysis of variance for principal component 2 in *CG17928* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	2.971	2.971	1.03	0.318
Sex	1	0.020	0.020	0.01	0.934
Age	1	2.950	2.950	1.02	0.320
Treatment*Sex	1	6.072	6.072	2.10	0.157
Treatment*Age	1	0.250	0.250	0.09	0.771
Error	34	98.342	2.892		
Total	39	110.606			

Table 5.34: Analysis of variance for principal component 3 in *CG17928* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	2.610	2.610	1.72	0.199
Sex	1	5.479	5.479	3.61	0.066
Age	1	0.037	0.037	0.02	0.877
Treatment*Sex	1	0.088	0.088	0.06	0.811
Treatment*Age	1	5.364	5.364	3.52	0.069
Error	34	51.630	1.519		
Total	39	65.208			

Table 5.35: Analysis of variance for principal component 4 in *CG17928* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.341	0.341	0.30	0.585
Sex	1	1.840	1.840	1.64	0.209
Age	1	1.403	1.403	1.25	0.271
Treatment*Sex	1	1.123	1.123	1.00	0.324
Treatment*Age	1	0.132	0.132	0.12	0.734
Error	34	38.142	1.122		
Total	39	42.983			

Table 5.36: Analysis of variance for principal component 5 in *CG17928* analysis

Cyt-b5-r

	PC1	PC2	PC3
Eigenvalue	14.367	6.470	1.269
% variance	55.3	24.9	4.9

Table 5.37: Eigenvalues and % variance explained for principal components in *Cyt-b5-r* analysis.

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.003	0.003	0.00	0.980
Sex	1	328.696	328.696	73.71	**<0.001**
Age	1	79.691	79.691	17.87	**<0.001**
Treatment*Sex	1	0.014	0.014	0.00	0.956
Treatment*Age	1	0.295	0.295	0.07	0.798
Error	34	151.618	4.459		
Total	39	560.317			

Table 5.38: Analysis of variance for principal component 1 in *Cyt-b5-r* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.028	0.028	0.02	0.877
Sex	1	35.731	35.731	31.58	**<0.001**
Age	1	177.275	177.275	156.70	**<0.001**
Treatment*Sex	1	0.411	0.411	0.36	0.551
Treatment*Age	1	0.408	0.408	0.36	0.552
Error	34	38.465	1.131		
Total	39	252.318			

Table 5.39: Analysis of variance for principal component 2 in *Cyt-b5-r* analysis

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.867	0.867	0.65	0.424
Sex	1	1.732	1.732	1.31	0.261
Age	1	1.668	1.668	1.26	0.270
Treatment*Sex	1	0.008	0.008	0.01	0.937
Treatment*Age	1	0.129	0.129	0.10	0.757
Error	34	45.079	1.326		
Total	39	49.483			

Table 5.40: Analysis of variance for principal component 3 in *Cyt-b5-r* analysis.

5.3.2 Analysis of unsaturated:saturated CHC ratios

ANOVA results for all analyses of unsaturated:saturated CHC ratios are shown in Tables 5.41 to 5.47. As above, significant P-values are shown in bold in the following tables, and "almost significant" P-values are in italics. The data fit approximately to a normal distribution. After correcting for multiple testing, Treatment was significant in the analyses for *desat2*, *CG8630* and *CG9743*, and borderline for *CG15531*. The interaction between Treatment and Sex was significant in the analyses for *desat2* and *CG9743*. The Treatment*Age interaction was significant in analyses of *CG8630* and *CG15531*. The Treatment*Sex interaction in the *CG9747* ANOVA had P = 0.033, less than 0.05 but non-significant using the cut-off of 0.007. Age and/or Sex were also significant in most of the analyses.

The least squares means estimates for unsaturated:saturated CHC ratio are shown in Figures 5.3 to 5.9. In the *desat2*, *CG8630* and *CG15531* experiments, the least squares means for RNAi males and females are lower than those of control males and females. The ANOVA results indicate that Treatment has had a significant effect on the ratios for *desat2* and *CG8630*. For *CG15531* it was "almost significant". For *CG9743*, *CG17928* and *Cyt-b5-r*, the means are higher in RNAi males and females than those of control males and females. Out of these, only *CG9743* has a significant P-value for Treatment in the ANOVA results. For *CG9747*, the RNAi males' mean is higher than that of control males, but RNAi females have a lower mean than control females. Treatment does not have a significant P-value in the ANOVA results for *CG9747*. These results are summarised in Table 5.48.

desat2

Source	DF	Seq. SS	MS	F	P
Treatment	1	46.333	46.333	39.91	**<0.001**
Sex	1	30.323	30.323	26.13	**<0.001**
Age	1	0.178	0.178	0.15	0.698
Treatment*Sex	1	10.431	10.431	8.99	**0.005**
Treatment*Age	1	0.588	0.588	0.51	0.482
Error	34	39.462	1.161		
Total	39	127.314			

Table 5.41: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *desat2* experiment.

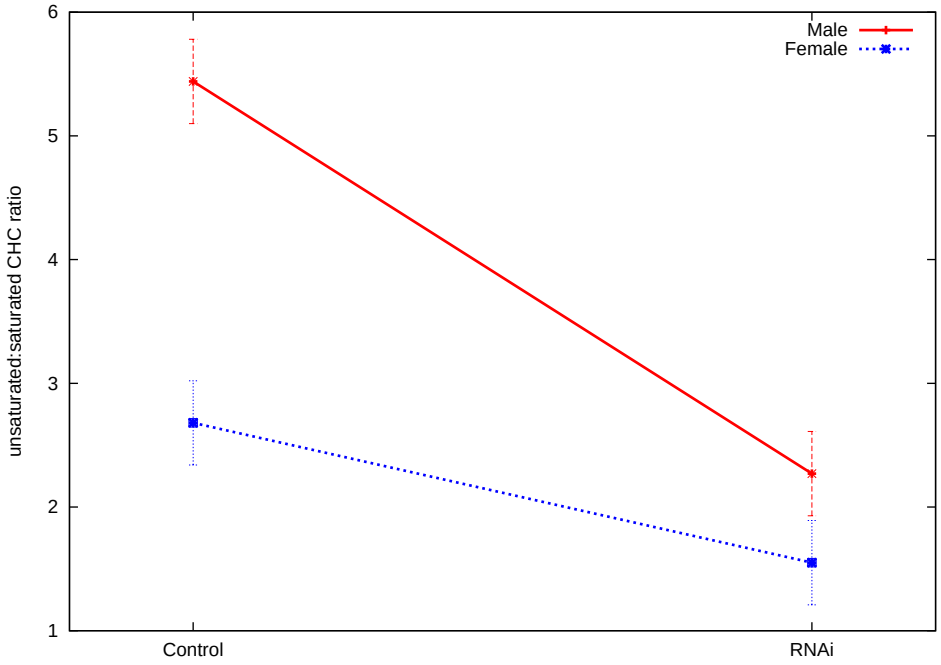


Figure 5.3: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *desat2* experiment

CG8630

Source	DF	Seq. SS	MS	F	P
Treatment	1	34.543	34.543	59.35	**<0.001**
Sex	1	37.620	37.620	64.64	**<0.001**
Age	1	3.342	3.342	5.72	*0.022*
Treatment*Sex	1	2.134	2.134	3.67	0.064
Treatment*Age	1	16.518	16.518	28.38	**<0.001**
Error	34	19.786	0.582		
Total	39	113.944			

Table 5.42: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *CG8630* experiment.

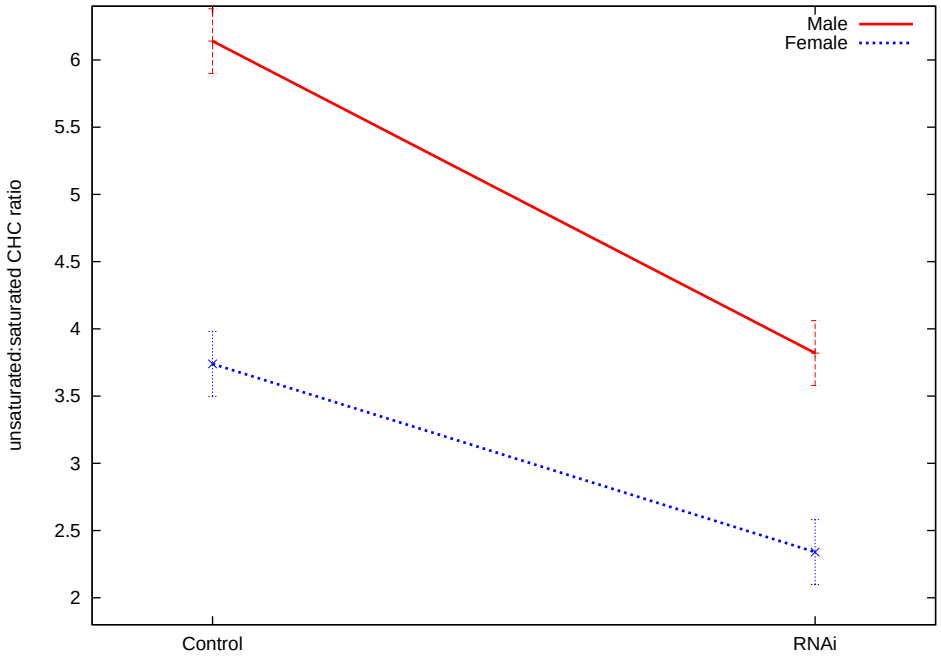


Figure 5.4: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *CG8630* experiment

CG9743

Source	DF	Seq. SS	MS	F	P
Treatment	1	13.932	13.932	14.95	**0.001**
Sex	1	63.161	63.161	67.73	**<0.001**
Age	1	41.299	41.299	44.29	**<0.001**
Treatment*Sex	1	11.958	11.958	12.82	**0.001**
Treatment*Age	1	1.337	1.337	1.43	0.239
Error	34	31.705	0.932		
Total	39	163.391			

Table 5.43: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *CG9743* experiment.

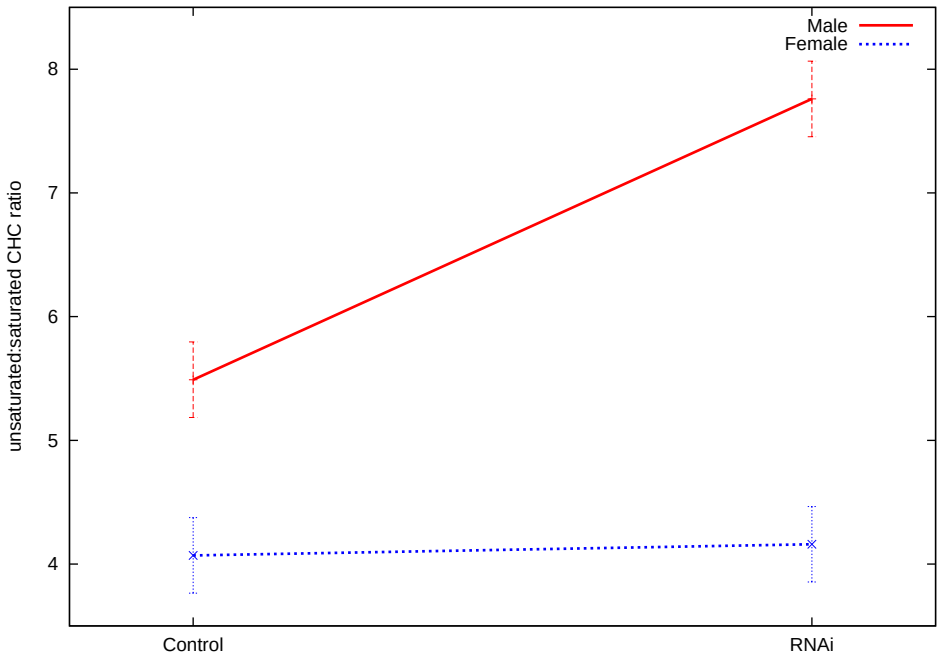


Figure 5.5: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *CG9743* experiment

CG9747

Source	DF	Seq. SS	MS	F	P
Treatment	1	0.8206	0.8206	1.21	0.278
Sex	1	28.0078	28.0078	41.44	**<0.001**
Age	1	0.8872	0.8872	1.31	0.260
Treatment*Sex	1	3.3267	3.3267	4.92	*0.033*
Treatment*Age	1	0.0234	0.0234	0.03	0.854
Error	34	22.9805	0.676		
Total	39	56.0462			

Table 5.44: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *CG9747* experiment.

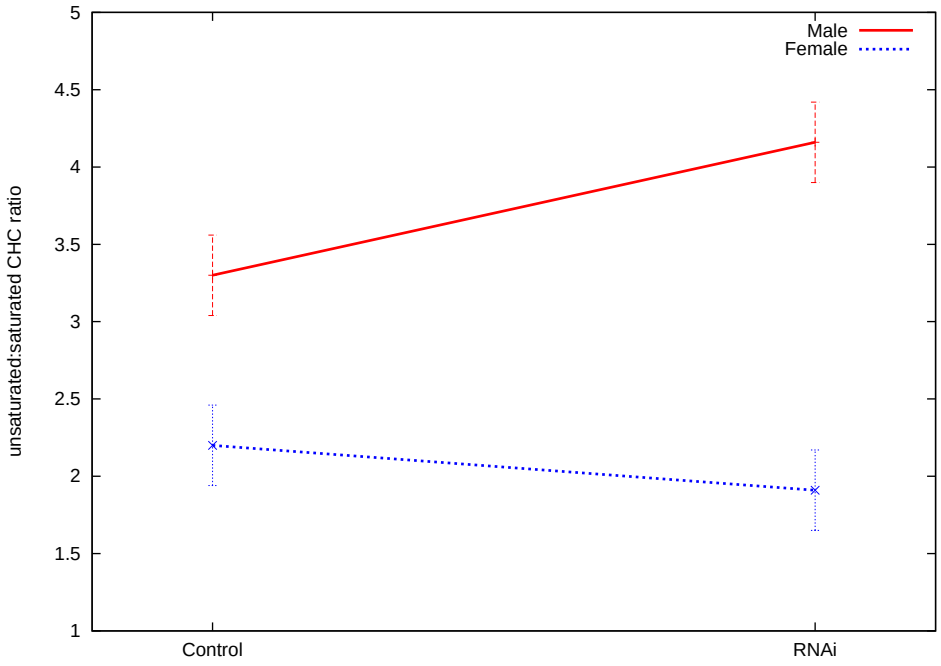


Figure 5.6: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *CG9747* experiment

CG15531

Source	DF	Seq. SS	MS	F	P
Treatment	1	20.736	20.736	8.20	*0.007*
Sex	1	196.550	196.550	77.74	**<0.001**
Age	1	9.695	9.695	3.83	0.058
Treatment*Sex	1	2.301	2.301	0.91	0.347
Treatment*Age	1	41.663	41.663	16.48	**<0.001**
Error	34	85.957	2.528		
Total	39	356.903			

Table 5.45: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *CG15531* experiment.

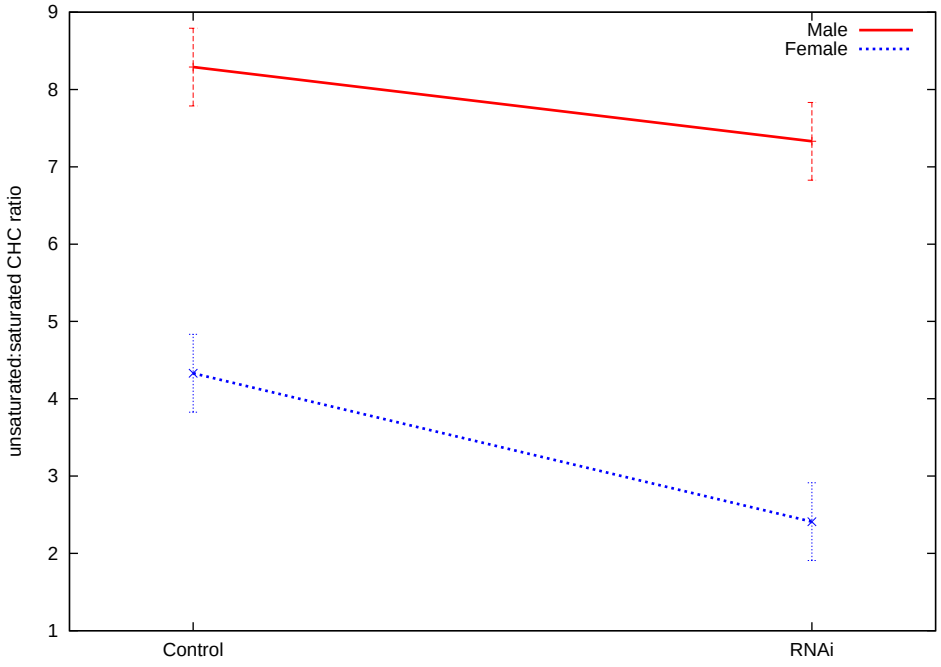


Figure 5.7: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *CG15531* experiment

CG17928

Source	DF	Seq. SS	MS	F	P
Treatment	1	4.275	4.275	1.36	0.252
Sex	1	31.031	31.031	9.88	**0.003**
Age	1	0.110	0.110	0.03	0.853
Treatment*Sex	1	3.055	3.055	0.97	0.331
Treatment*Age	1	0.270	0.270	0.09	0.771
Error	34	106.778	3.141		
Total	39	145.518			

Table 5.46: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *CG17928* experiment.

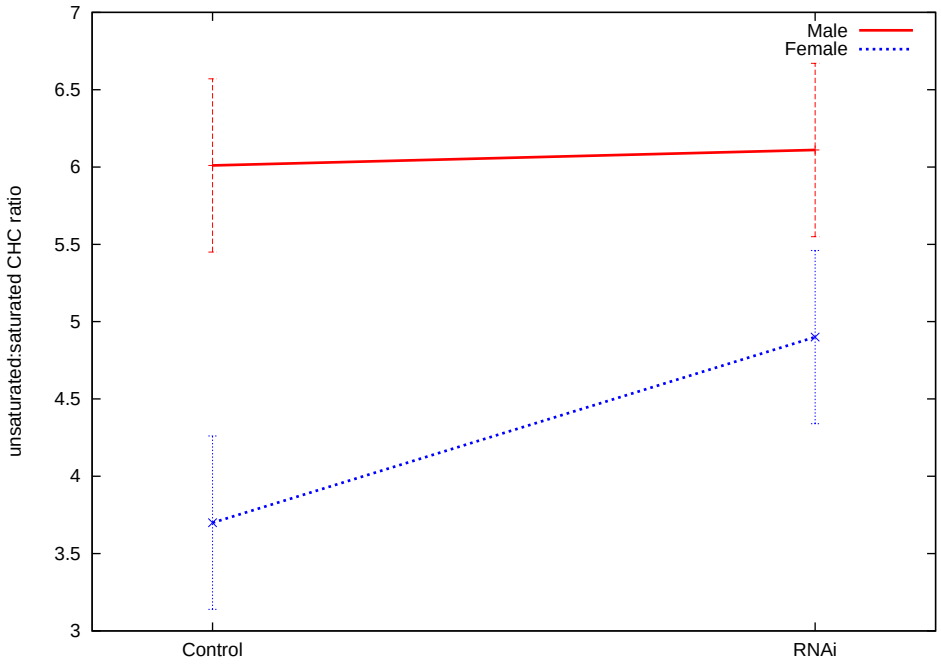


Figure 5.8: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *CG17928* experiment

Cyt-b5-r

Source	DF	Seq. SS	MS	F	P
Treatment	1	4.031	4.031	5.69	*0.023*
Sex	1	29.415	29.415	41.51	**<0.001**
Age	1	59.360	59.360	83.77	**<0.001**
Treatment*Sex	1	0.010	0.010	0.01	0.905
Treatment*Age	1	0.035	0.035	0.05	0.825
Error	34	24.092	0.709		
Total	39	116.943			

Table 5.47: Analysis of variance results for the ratios of unsaturated:saturated CHCs in *Cyt-b5-r* experiment.

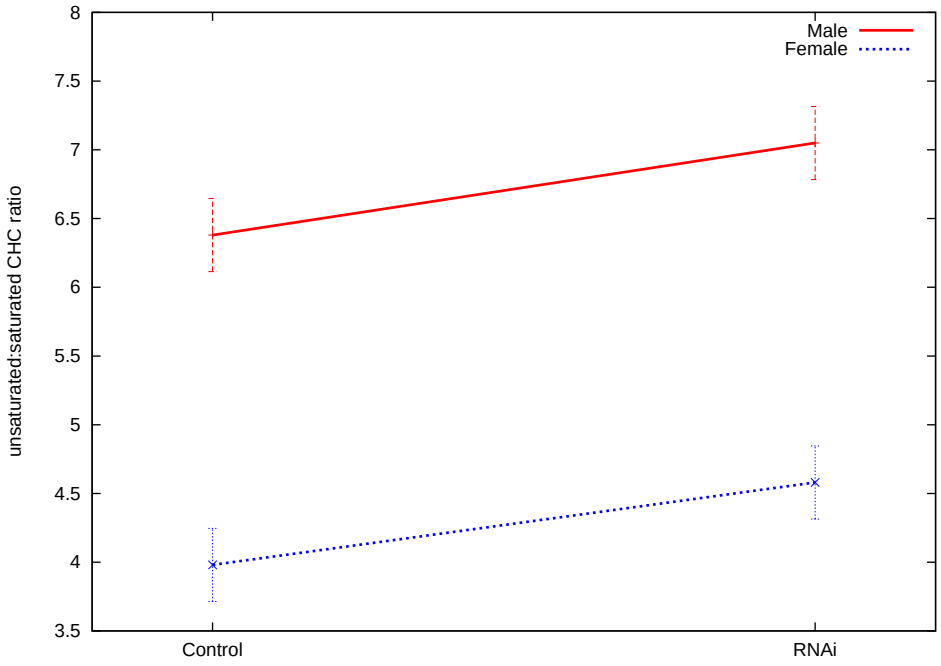


Figure 5.9: Least squares means estimates and standard deviations of unsaturated:saturated CHC ratios in Control and RNAi males and females from *Cyt-b5-r* experiment

Summary of analyses of unsaturated:saturated CHC ratios

Gene	Significant term(s)	Change in ratio
desat2	Treatment	Down in RNAi ♂and ♀
	Sex	
	Treatment*Sex	
CG8630	Treatment	Down in RNAi ♂and ♀
	Sex	
	Treatment*Age	
CG9743	Treatment	Up in RNAi ♂
	Sex	Unchanged in ♀
	Age	
	Treatment*Sex	
CG9747	Sex	Up in RNAi ♂
		Down in RNAi ♀
CG15531	Sex	Down in RNAi ♂and ♀
	Treatment*Age	
CG17928	Sex	Unchanged in ♂
		Up in RNAi ♀
Cyt-b5-r	Sex	Up in RNAi ♂and ♀
	Age	

Table 5.48: Summary of analyses of unsaturated:saturated CHC ratios.

5.4 Discussion

5.4.1 Principal components analysis

Principal components analysis (PCA) found that both age and sex had significant effects in all knock-down experiments. It is of course widely known that mature *D. melanogaster* have sexually dimorphic CHCs (Antony and Jallon, 1982), so the effect of sex is not unexpected. Neither is that of age: it is also known that the CHC profiles of both male and female flies change over time. Immature *D. melanogaster* male and female adults both produce much longer-chain CHCs than mature adults; in both sexes immature males and females also possess dienic CHCs, not found in mature adult males. This is thought to be the reason young males are sometimes courted by mature males (Antony and Jallon, 1982; Ferveur, 2005; Pechine *et al.*, 1988).

Significant effect of RNAi knock-down was detected in principal component 5 in the *CG8630* analysis, and the interaction between Treatment and Age is also significant here. The Treatment*Sex interaction was significant in principal component 4 for *CG15531*. According to FlyAtlas, the online database of *D. melanogaster* adult gene expression, *CG8630* is down-regulated in the tissue category "Adult carcass", which includes the oenocytes (Chintapalli *et al.*, 2007). It is perhaps surprising, therefore, that attempted RNAi knock-down of this gene specifically in the oenocytes would have an effect on CHC profile. The data in FlyAtlas is from 7-day-old *D. melanogaster* Canton-S flies. Given that the CHC profile of younger flies differs from that of older ones, it is possible that certain desaturases may be expressed in younger flies and not in older ones. The flies used in the work presented here were aged 1-day-old and 4-days-old, so if *CG8630* is expressed early and then down-regulated, this could explain the effect detected in PCA. FlyAtlas shows that *CG15531* is up-regulated in "Adult carcass" in 7-day-old flies, so there is potential for it to affect CHCs.

Two other genes are up-regulated in "Adult carcass" according to FlyAtlas, these are *CG17928* and *Cyt-b5-r*. Although they are expressed in this tissue category, it is possible that they are involved in fatty acid metabolism in other biological

processes and are not required for CHC modification. This would explain the lack of significance detected in these genes. A few genes have "almost significant" P-values (in italics in Tables 5.2 to 5.40). The sample size in this study was fairly small: 5 flies per sex per treatment. A larger sample size would give greater power to detect real effects, and it is possible that some of these "almost significant" P-values would become significant. Of course, there is also potential for the opposite result, that the significant or nearly significant results seen here are due to sampling error. Larger sample sizes would provide a more accurate picture of the effects of RNAi knock-down of these genes.

In general, the data used in the PCA were not normally distributed, and so the linear model may not fit adequately. Because of this, the results of this analysis may not be 100% reliable. The reason for this is likely to be because the CHC data contains many zero values. The main factors contributing to this are dienic CHCs, which were only present in females, and mostly only in 4-day-old females, while males and most 1-day-old females did not have any dienes. For these columns, the variance in males is of course zero, given they all have the same value (zero) for any particular diene. This produces skewed data because the females have at least some variance in these columns. In other analyses it may be possible to eliminate zero values, for example by combining certain columns or by simply removing the columns or rows containing zeros. However these approaches were not deemed sensible in this analysis: given that dienic CHCs are some of the most important when it comes to *Drosophila* pheromonal communication, and that they are directly affected by certain desaturases, removing them would remove a source of potentially interesting variation. The diene columns were therefore kept in the data. However, the results should be taken with caution.

5.4.2 Analysis of unsaturated:saturated CHC ratios

The ratios of unsaturated:saturated CHCs were more normally distributed, and so fit the linear models better. Treatment appears to have lowered the ratio in the *desat2*, *CG8630* and *CG15531* experiments, though the P-value for *CG15531* is equal to the cut-off of 0.007. This more or less concurs with what was found in

the PCA analysis. There has also been a significant interaction between Treatment and Sex in *desat2*, and from the least squares means it appears that Treatment has lowered the ratio more in males than in females (Figure 5.3). *desat2* is known to be involved in CHC modification in certain strains of *D. melanogaster*. It has been shown that *desat2* is not expressed in Canton-S and other Cosmopolitan strains, or else is only expressed at a very low level, due to a 16bp deletion in the putative promoter region (Dallerac *et al.*, 2000; Michalak *et al.*, 2007; Takahashi *et al.*, 2001). The RNAi flies used were created using the w^{1118} strain which has a Canton-S genetic background (Dietzl *et al.*, 2007). It is perhaps surprising therefore that attempted RNAi knock-down of this gene should have an effect. There is a possibility for non-specific targetting by the RNAi system: *desat2* is very similar in sequence to *desat1*, and so it is possible that *desat1* is also being targeted by the RNAi machinery. This would cause a reduction in the amount of unsaturated CHCs and a corresponding rise in the amount of linear ones (Wicker-Thomas *et al.*, 2009), and hence a lower unsaturated:saturated CHC ratio.

This analysis also found that Treatment had a significant effect in the *CG9743* analysis, apparently causing an increase in the unsaturated:saturated CHC ratio, particularly in males. This was unexpected: the function of a desaturase is to create unsaturations, and so removing its function should result in less unsaturations, if the enzyme is involved in CHC modification. It is possible that this enzyme has some other function in controlling CHC production but it is not clear what that might be from these results. It is important to bear in mind that the desaturases act in a complex network with many other enzymes; perhaps the reduction in activity of *CG9747*, directly or indirectly, leads to an increase in activity of some other desaturase. The other possible explanation is that this is an artefact of the small sample sizes. No significant effect of Treatment was detected for *CG17928* or *Cyt-b5-r*. This indicates that these genes do not play a role in CHC production.

5.4.3 Concluding remarks

The results of the PCA and the analysis of unsaturated:saturated CHC ratios suggest that, apart from the well-studied functions *desat1*, *desat2* and *desatF*, other

desaturases may have roles to play in the production of CHCs in *D. melanogaster*. The caveat of small sample size applies, but with a larger sample size it would be possible to obtain a more definitive answer to the question of which desaturases are important for this process. It would also be of interest to investigate the expression pattern of *CG8630* further, to discover whether it is expressed in the oenocytes in younger adults.

Aside from the small sample size, another caveat for these analyses was that the RNAi knock-down was not verified by RNA extraction and Q-PCR. The *GAL4* driver used was verified to function in the oenocytes and successfully drove RNAi of other desaturases in these cells by Wicker-Thomas *et al.* (2009). However, it would strengthen this study if it could be said that each gene was found to have been knocked down by investigating the amount of mRNA present. This process is difficult for the oenocytes, as they form very small amounts of tissue and this would need to be dissected specifically and quickly enough to prevent RNA degradation. This was not possible in the scope of this project, however it is concievable that it would be possible in future.

Bibliography

Antony, C. and Jallon, J. M. (1982). The chemical basis for sex recognition in *Drosophila melanogaster*. *Journal of Insect Physiology*, **28**:873–880.

Chintapalli, V. R., Wang, J. and Dow, J. A. T. (2007). Using FlyAtlas to identify better *Drosophila* models of human disease. *Nature Genetics*, **39**:715–720.

Cronin, S. J. F., Nehme, N. T. *et al.* (2009). Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. *Science*, **325**:340–343.

Dallerac, R., Labeur, C., Jallon, J.-M., Knipple, D. C., Roelofs, W. L. and Wicker-Thomas, C. (2000). A Δ9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **97**:9449–9454.

Dietzl, G., Chen, D. *et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, **448**:151–156.

Duffy, J. B. (2002). GAL4 system in *Drosophila*: a fly geneticist's swiss army knife. *Genesis*, **34**:1–15.

Ferveur, J. F. (2005). Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behavior Genetics*, **35**:269–295.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**:806–811.

Hannon, G. J. (2002). RNA interference. *Nature*, **418**:244–251.

Kennerdell, J. R. and Carthew, R. W. (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnology*, **17**:896–898.

Kitson, F. G., Larsen, B. S. and McEwan, C. N. (1996). *Gas chromatography and mass spectrometry: a practical guide*. San Diego: Academic Press.

Maeda, I., Kohara, Y., Yamamoto, M. and Sugimoto, A. (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Current Biology*, **11**:171–176.

Michalak, P., Malone, J. H., Lee, I. T., Hoshino, D. and Ma, D. (2007). Gene expression polymorphism in *Drosophila* populations. *Molecular Ecology*, **16**:1179–1189.

Mummery-Widmer, J. L., Yamazaki, M. *et al.* (2009). Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. *Nature*, **458**:987–992.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990). lntroduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. *The Plant Cell*, **2**:279–289.

Pechine, J. M., Antony, C. and Jallon, J. M. (1988). Precise characterization of cuticular hydrocarbons in young *Drosophila* by mass spectrometry. *Journal of Chemical Ecology*, **14**:1071–1085.

Perrimon, N., Ni, J. Q. and Perkins, L. (2010). *In vivo* RNAi: today and tomorrow. *Cold Spring Harbor Perspectives in Biology*, **2**:a003640.

Roman, G. (2004). The genetics of *Drosophila* transgenics. *Bioessays*, **26**:1243–1253.

Sönnichsen, B., Koski, L. B. *et al.* (2005). Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature*, **434**:462–469.

Takahashi, A., Tsaur, S.-C., Coyne, J. A. and Wu, C.-I. (2001). The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA*, **98**:3920–3925.

Terenius, O., Papanicolaou, A. *et al.* (2011). RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. *Journal of Insect Physiology*, **57**:231–245.

van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N. M. and Stuitje, A. R. (1990). Flavonoid genes in *Petunia*: Addition of a limited number of gene copies may lead to a suppression of gene expression. *The Plant Cell*, **2**:291–299.

Vance, V. and Vaucheret, H. (2001). RNA silencing in plants – defense and counterdefense. *Science*, **292**:2277–2280.

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends in Genetics*, **17**:449–459.

Wicker-Thomas, C., Guenachi, I. and Keita, Y. F. (2009). Contribution of oenocytes and pheromones to courtship behaviour in *Drosophila*. *BMC Biochemistry*, **10**:21.

Wilkins, C., Dishongh, R., Moore, S. C., Whitt, M. A., Chow, M. and Machaca, K. (2005). RNA interference is an antiviral defence mechanism in *Caenorhabditis elegans*. *Nature*, **436**:1044–1047.

Chapter 6

Discussion

The release of the complete genome sequences of 12 *Drosophila* species has provided researchers with the opportunity for comparative analyses across a large amount of evolutionary time. Following the publication of the genome sequences came many studies examining gene family evolution across the phylogeny. Hahn *et al.* (2007) used this newly available resource to study gain and loss of loci across the whole genome. They found large numbers of gains and losses, with some instances of extreme expansion or contraction in gene family size, particularly in sex-related genes. They note that although the overall number of genes has remained relatively constant over time, there has been rapid turnover in gain and loss of individual genes and families (Hahn *et al.*, 2007).

Studies of individual gene families so far tend to find that they are generally governed by purifying selection, with more relaxed constraints and sometimes positive selection on duplicated genes. This was true in a study of the odorant-binding protein (OBP) family, a group of proteins which establish the initial contact with odorant molecules (Vieira *et al.*, 2007). This analysis identified 595 genes, with 43 gains and 28 losses, 13 of which were pseudogenisations. The authors note that evolution in this family was more rapid in the specialists *D. sechellia* and *D. erecta* than in their closely related generalist species *D. simulans* and *D. yakuba*, suggesting that the OBP family was involved in the process of specialisation (Vieira *et al.*, 2007).

Guo and Kim (2007) performed a comparative analysis of olfactory receptors,

identifying 59 orthologous groups of genes and noting a large amount of duplication and loss, particularly in *D. willistoni*, *D. grimshawi*, *D. pseudoobscura* and *D. persimilis*. In a similar study of this gene family, Gardiner *et al.* (2008) analysed olfactory and gustatory receptors and found that gene family size varied considerably between different species, with many genes having undergone pseudogenisation, and many lineage-specific duplications. They also note that the majority of the loci were under strong purifying selection, although this is relaxed in genes that have been duplicated.

The desaturase gene family also appears to be under strong purifying selection for the most part, with all instances of positive selection being in duplicated genes. No pseudogenisation was found among the desaturases, however, indicating that they are under stronger constraints, perhaps due to the fact that there has been less duplication in this family – most desaturases are only present in a single copy. The size of the desaturase family remains fairly consistent between different species: only *desatF* and *desat2* appear to have been lost, and although *desatF* has been duplicated several times, at most a single species has three copies. Recently, a similar comparative genomic study of the *Drosophila* desaturases was published, which combined homology-based searching with synteny analysis to identify orthologous genes (Fang *et al.*, 2009). Their analysis identified most of the loci found in this thesis, excluding two duplicated genes in the form of *desat1b* in *D. pseudoobscura* and *D. persimilis*, and a duplicate of *desat2* in *D. ananassae*. They also found evidence for the desaturases being generally governed by purifying selection, with relaxation or positive selection on branches following duplication, which concurs with this study.

The work presented in this thesis has identified novel duplicates in the *Drosophila* desaturase gene family. Two duplicates show strong evidence of being under positive selection: *desat1b* and *desatF*α in *D. pseudoobscura* and *D. persimilis*. It is shown that *desat1b* is expressed in both species, lending support to the idea that it is a functional gene, and not a pseudogene and therefore bolstering the claim of positive selection. Furthermore, it is expressed in a sex-biased manner and also demonstrates alternative splicing in its 5′ UTR at least in males of *D. pseudoobscura*. The sex-biased expression pattern in a gene likely to influence pheromones is particularly

interesting. The RNAi work in *D. melanogaster* has identified genes, which had not previously been associated with CHC production, as having a possible effect on it. These results demonstrate the potential of comparative genomic studies in furthering our understanding of gene family evolution and in discovering new candidates for genes that influence reproductive isolation.

Future directions

Much of the work presented here raises new questions to be answered. In Chapter 3, two genes were found to show evidence of positive selection (Tables 3.4 and 3.6). The *desat1b* gene of *D. pseudoobscura* and *D. persimilis* was subjected to further analysis; the other gene – *D. pseudoobscura* and *D. persimilis desatFα* – showed strong evidence for positive selection on the branch preceeding speciation, but after gene duplication (Figure 3.5). The regulation of the *desatF* gene is known to have undergone rapid evolution in many species (Shirangi *et al.*, 2009). It would be interesting to investigate the duplicates in the *obscura* group species further to see what their patterns of expression are, and how the positive selection detected in Chapter 3 has affected the function of the gene.

There are many possibilities to follow up the work so far on *desat1b* in *D. pseudoobscura* and *D. persimilis* (Chapter 4). These species naturally occur in North America; *D. persimilis* inhabits only the western USA, while *D. pseudoobscura* is found all across the country. There are therefore populations of *D. pseudoobscura* in the western US that are sympatric with *D. persimilis*, and those found further east that are allopatric. These different populations of *D. pseudoobscura* have been found to show different levels of behavioural isolation from *D. persimilis*, with females of sympatric *D. pseudoobscura* being less willing to mate with *D. persimilis* males than allopatric females are (Noor, 1995). *desat1b* is located on the X chromosome in *D. pseudoobscura*, which has been implicated in courtship in these species and strains (e.g. Noor *et al.*, 2001; Ortiz-Barrientos *et al.*, 2004, but see Barnwell and Noor 2008). It would be interesting to see whether the expression of *desat1b* differs between sympatric and allopatric strains of *D. pseudoobscura*, and could be contributing to this difference in isolation levels. This could be done using RT-PCR,

with primers designed based on the sequence of the mRNA found in Chapter 4 using RACE. Q-PCR could also be used here, to measure the levels of expression in males and females of the different strains, and also in different tissues to determine where in the fly the gene is expressed. An oenocyte-specific expression pattern would of course suggest involvement in CHC production. It would also be of interest to determine whether both sexes of both species and/or strains of *D. pseudoobscura* demonstrate alternative splicing in the 5′ UTR, as was found in *D. pseudoobscura* males.

Following on from expression analysis, it is important to find out the function of the protein product of *desat1b*. Given that it is so closely related to *desat1*, there is a possibility it also has a part to play in CHC modification. However, this is merely an assumption, and must be investigated empirically. Although it is not possible yet to buy ready-made RNAi lines of *D. pseudoobscura* or *D. persimilis*, as it is for *D. melanogaster*, RNAi is increasingly being performed in non-model organisms. It should be possible, therefore, to create lines of *D. pseudoobscura* and *D. persimilis* carrying an RNAi hairpin construct, which could then be used to knock down *desat1b* expression. To determine whether the protein is even translated is another possible avenue for investigation. Methods such as Western blotting can detect specific proteins in a protein extract from tissue or a whole organism. Since most of the positive selection was localised at the N-terminal region, it would be of interest to determine the function of this region and of the positively-selected residues. This could be investigated *in vitro* using site-directed mutagenesis, changing these residues and observing the effects, as has been done previously to determine the function of positively-selected sites (Bielawski *et al.*, 2004; Ilvarsson *et al.*, 2003; Norrgård *et al.*, 2006; Sawyer *et al.*, 2005).

It is as yet unknown whether any other species in the *obscura* group possess this gene: this would be relatively easy to determine, using the primers already designed in this thesis and attempting to amplify the gene from DNA extracted from other species. A caveat to all work involving amplification or probing of this gene using complementary oligonucleotides must be to ensure that only *desat1b* is targeted by the oligo, and not its close relative, *desat1*, since they are very similar in sequence.

The RNAi work in Chapter 5 also produced some potentially interesting results, particularly with the genes *CG8630* and *CG9743* having significant effects on the CHC profile. These are genes that up to now were unlinked to CHC modification, so it is of particular interest to follow this analysis up and determine what their exact function is. An obvious next step here is to repeat the RNAi experiments using a much larger sample size, for example 20 flies per category as opposed to 5, which was used here. This would provide much more statistical power to detect real effects in the data. Also potentially important to look at would be the mRNA levels in RNAi and Control flies – this was not done here in part because the *GAL4* driver was already verified to function as expected, but a Q-PCR analysis would strengthen the study and any conclusions drawn. RNAi experiments on other genes involved in CHC production would also be interesting, for example the elongases, which act after desaturases to lengthen the fatty acid hydrocarbon chain and produce the mature CHC.

Bibliography

Barnwell, C. V. and Noor, M. A. F. (2008). Failure to replicate two mate preference QTLs across multiple strains of *Drosophila pseudoobscura*. *Journal of Heredity*, **99**:653–656.

Bielawski, J. P., Dunn, K. A., Sabehi, G. and Beja, O. (2004). Darwinian adaptations of proteorhodopsin to different light intensities in the marine environment. *Proceedings of the National Academy of Sciences of the USA*, **101**:14824–14829.

Fang, S., Ting, C.-T., Lee, C.-R., Chu, K.-H., Wang, C.-C. and Tsaur, S.-C. (2009). Molecular evolution and functional diversification of fatty acid desaturases after recurrent gene duplication in *Drosophila*. *Molecular Biology and Evolution*, **26**:1447–1456.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C. and Ritchie, M. G. (2008). *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Molecular Ecology*, **17**:1648–1657.

Guo, S. and Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular Biology and Evolution*, **24**:1198–1207.

Hahn, M. W., Han, M. V. and Han, S. G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, **3**:e197.

Ilvarsson, Y., Mackey, A. J., Edalat, M., Pearson, W. R. and Mannervik, B. (2003). Identification of residues in glutathione transferase capable of driving functional diversification in evolution. *Journal of Biological Chemistry*, **278**:8733–8738.

Noor, M. A. (1995). Speciation driven by natural selection in *Drosophila*. *Nature*, **375**:674–675.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., Almendarez, Y., Reilan, J. and Smith, K. R. (2001). The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution*, **55**:512–521.

Norrgård, M. A., Ivarsson, Y., Tars, K. and Mannervik, B. (2006). Alternative mutations of a positively selected residue elicit gain or loss of functionalities in enzyme evolution. *Proceedings of the National Academy of Sciences of the USA*, **103**:4876–4881.

Ortiz-Barrientos, D., Counterman, B. A. and Noor, M. A. F. (2004). The genetics of speciation by reinforcement. *PLoS Biology*, **2**:e416.

Sawyer, S. L., Wu, L. I., Emerman, M. and Malik, H. S. (2005). Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the USA*, **102**:2832–2837.

Shirangi, T. R., Dufour, H. D., Williams, T. M. and Carroll, S. B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biology*, **7**:e1000168.

Vieira, F. G., Sánchez-Garcia, A. and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology*, **8**:R235.

Appendices

Appendix A

Laboratory Protocols

A.1 Single fly DNA prep

This protocol is adapted from Gloor *et al.* (1991).

Ingredients for squishing buffer

1 ml	1M TrisHCl
0.5 ml	0.2M EDTA
0.5 ml	5M NaCl
to 100ml	H_2O
20 mg/ml	proteinase K

Method

1. Place one fly in a 1.5 ml microcentrifuge tube, and leave at -20°C.
2. Make 100 ml squishing buffer using all ingredients shown apart from proteinase K.
3. Decant 990 μl squishing buffer into a microcentrifuge tube, and add 10 μl 20mg/ml proteinase K, giving a final concentration of 200 μg/ml proteinase K.

4. Remove fly from -20°C.

5. Aspirate 50 μl squishing buffer with proteinase K and, without expelling the buffer, mash fly with the end of the tip.

6. Expel the buffer into the tube.

7. Incubate at 37°C for 30 minutes.

8. Incubate at 95°C for 2 minutes to inactivate proteinase K.

9. Store at -20°C for up to 30 days.

A.2 PCR

In a 1.5 ml microcentrifuge tube, make mastermix containing all reagents given in Table A.1, except Taq DNA polymerase. Multiply all amounts by the number of reactions (including one MilliQ H_2O negative control) plus one extra for every ten reactions, to ensure sufficient mastermix is made for all reactions. Vortex primers and dNTPs prior to use.

Reagent	Amount (μl) for 1 reaction
MilliQ H_2O	19.55
Forward Primer (50 pmol/µl)	0.15
Reverse Primer (50 pmol/µl)	0.15
10X NH_4	2.5
$MgCl_2$	0.75
dNTPs (10 nmol/µl)	0.8
Taq DNA Polymerase	0.1

Table A.1: Ingredients for a single 25 μl PCR reaction.

Add 1 μl template to separate 0.2 ml PCR tubes. Add the Taq DNA polymerase to the mastermix and vortex briefly to mix. Decant 24 μl mastermix into each PCR tube, making sure that there are no air bubbles separating the mastermix and template. Place tubes in a thermocycler and run the following program:

Step 1. 94°C for 3 minutes

Step 2. 94°C for 30 seconds

Step 3. Annealing temperature for 30 seconds

Step 4. 72°C for 1 minute

Step 5. Go to Step 2 29 more times

Step 6. 72°C for 5 minutes

Check all PCR reactions on a 2% agarose gel.

A.3 Agarose gel electrophoresis

2% Agarose Gel

Place combs and barriers to gel tray. Add 50 ml 0.5X TBE to 1 g agarose in a conical flask. Cover flask with cling film, pierce film, microwave until all agarose has dissolved. Allow gel to cool until it is not steaming, then add 2.5 μl ethidium bromide and swirl to mix. Pour gel into tray and allow to set. When gel has set, remove combs and barriers and pour over 50 ml 0.5X TBE.

Electrophoresis of PCR products

On a clear slide, pipette one 2 μl drop of loading dye for each PCR reaction. To each drop, add 5 μl PCR reaction. Load these on the gel, as well as 2.5 μl GeneRuler 100 bp ladder. Connect the power supply and run at 90 volts for 20 minutes. After the run, examine gel under UV light to visualise PCR products.

A.4 Corn meal fly food

Put all the ingredients in Table A.2 into a pot. Add the water and stir continuously. Once it boils, reduce heat and simmer while stirring constantly for 15 minutes. Do not allow to burn. Cool for 10-15 minutes.

Ingredient	Amount
Corn meal	50 g
Brewer's yeast	50 g
Glucose	80 g
Agar	10 g
H_2O	1 L

Table A.2: Ingredients for 1 L corn meal fly food.

While to cornmeal mixture is cooling, mix together 7 g baker's yeast and 40 ml H_2O and boil. Add directly to the cooled cornmeal mixture, as well as 27 ml 10% nipagin.

Add cornmeal mixture to empty vials. Cover vials with a cloth and leave to cool overnight, then add foam lids to all vials and refrigerate.

Appendix B

PAML Control Files

B.1 Site-based models

B.1.1 M7 and M8 models

```
                seqfile = desat1_allsp_codons_4paml.phy
                      * sequence data file name
     treefile = species.tree   * tree structure file name

      outfile = desat1_M7M8.paml          * main result
         file name
        noisy = 3   * 0,1,2,3,9: how much rubbish on the
           screen
      verbose = 0   * 1: detailed output, 0: concise
         output
      runmode = 0   * 0: user tree;  1: semi-automatic;
         2: automatic
                    * 3: StepwiseAddition; (4,5):
                       PerturbationNNI; -2: pairwise

       seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
     CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
```

```
      codon table
       clock = 0   * 0: no clock, unrooted tree, 1:
          clock, rooted tree
       aaDist = 0   * 0:equal, +:geometric; -:linear,
         {1-5:G1974,Miyata,c,p,v}
       model = 0

     NSsites = 7 8
                  * 0:one w; 1:NearlyNeutral; 2:
                     PositiveSelection; 3:discrete;
                  * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                     beta&w;9:beta&gamma;10:3normal
       icode = 0   * 0:standard genetic code; 1:
          mammalian mt; 2-10:see below
       Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
           4:all

   fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
      estimated
       kappa = .3   * initial or fixed kappa
   fix_omega = 0   * 1: omega or omega_1 fixed, 0:
      estimate
       omega = 1.3  * initial or fixed omega, for codons
           or codon-based AAs
       ncatG = 10   * # of categories in the dG or AdG
          models of rates

       getSE = 0   * 0: don't want them, 1: want S.E.s
          of estimates
RateAncestor = 0   * (0,1,2): rates (alpha>0) or
   ancestral states (1 or 2)
```

```
Small_Diff = .45e-6
 cleandata = 1  * remove sites with ambiguity data (1:
    yes, 0:no)?
fix_blength = 0  * 0: ignore, -1: random, 1: initial,
   2: fixed
```

B.1.2 M8a model

```
    seqfile = desat1_allsp_codons_4paml.phy    *
       sequence data file name
   treefile = species.tree   * tree structure file name

    outfile = desat1_M8a.paml          * main result
       file name
      noisy = 3   * 0,1,2,3,9: how much rubbish on the
         screen
    verbose = 0   * 1: detailed output, 0: concise
       output
    runmode = 0   * 0: user tree;  1: semi-automatic;
       2: automatic
                  * 3: StepwiseAddition; (4,5):
                     PerturbationNNI; -2: pairwise

     seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
   CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
      codon table
       clock = 0   * 0: no clock, unrooted tree, 1:
          clock, rooted tree
      aaDist = 0   * 0:equal, +:geometric; -:linear,
         {1-5:G1974,Miyata,c,p,v}
       model = 0
```

```
      NSsites = 8
                    * 0:one w; 1:NearlyNeutral; 2:
                       PositiveSelection; 3:discrete;
                    * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                       beta&w;9:beta&gamma;10:3normal
        icode = 0   * 0:standard genetic code; 1:
           mammalian mt; 2-10:see below
        Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
            4:all

    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
       estimated
        kappa = .3   * initial or fixed kappa
    fix_omega = 1   * 1: omega or omega_1 fixed, 0:
       estimate
        omega = 1  * initial or fixed omega, for codons
           or codon-based AAs
        ncatG = 10   * # of categories in the dG or AdG
           models of rates

        getSE = 0   * 0: don't want them, 1: want S.E.s
           of estimates
  RateAncestor = 0   * (0,1,2): rates (alpha>0) or
     ancestral states (1 or 2)

    Small_Diff = .45e-6
     cleandata = 1  * remove sites with ambiguity data (1:
       yes, 0:no)?
   fix_blength = 0  * 0: ignore, -1: random, 1: initial,
     2: fixed
```

B.2 Branch models

B.2.1 One-ratio model

```
        seqfile = desat1_all_codons_4paml.phy    *
           sequence data file name
     treefile = one_ratio.tree   * tree structure file
       name

      outfile = one_ratio.paml          * main result
        file name
       noisy = 3   * 0,1,2,3,9: how much rubbish on the
         screen
      verbose = 0   * 1: detailed output, 0: concise
        output
      runmode = 0   * 0: user tree;  1: semi-automatic;
        2: automatic
                    * 3: StepwiseAddition; (4,5):
                       PerturbationNNI; -2: pairwise

      seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
       codon table
       clock = 0   * 0: no clock, unrooted tree, 1:
          clock, rooted tree
       aaDist = 0   * 0:equal, +:geometric; -:linear,
         {1-5:G1974,Miyata,c,p,v}
       model = 0

      NSsites = 0
                    * 0:one w; 1:NearlyNeutral; 2:
                       PositiveSelection; 3:discrete;
```

```
                    * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                       beta&w;9:beta&gamma;10:3normal
        icode = 0   * 0:standard genetic code; 1:
           mammalian mt; 2-10:see below
        Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
            4:all

    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
       estimated
        kappa = .3   * initial or fixed kappa
    fix_omega = 0   * 1: omega or omega_1 fixed, 0:
       estimate
        omega = 1.3  * initial or fixed omega, for codons
            or codon-based AAs
        ncatG = 10   * # of categories in the dG or AdG
           models of rates

        getSE = 0   * 0: don't want them, 1: want S.E.s
           of estimates
 RateAncestor = 0   * (0,1,2): rates (alpha>0) or
    ancestral states (1 or 2)

   Small_Diff = .45e-6
    cleandata = 1  * remove sites with ambiguity data (1:
       yes, 0:no)?
  fix_blength = 0  * 0: ignore, -1: random, 1: initial,
     2: fixed
```

B.2.2 Two and three-ratio models

```
        seqfile = desat1_all_codons_4paml.phy    *
           sequence data file name
```

```
      treefile = two_ratios.tree   * tree structure file
        name

       outfile = two_ratios_ratios.paml          * main
         result file name
        noisy = 3   * 0,1,2,3,9: how much rubbish on the
           screen
      verbose = 0   * 1: detailed output, 0: concise
         output
      runmode = 0   * 0: user tree;  1: semi-automatic;
         2: automatic
                    * 3: StepwiseAddition; (4,5):
                       PerturbationNNI; -2: pairwise

      seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
       codon table
        clock = 0   * 0: no clock, unrooted tree, 1:
           clock, rooted tree
       aaDist = 0   * 0:equal, +:geometric; -:linear,
          {1-5:G1974,Miyata,c,p,v}
        model = 2

      NSsites = 0
                    * 0:one w; 1:NearlyNeutral; 2:
                       PositiveSelection; 3:discrete;
                    * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                       beta&w;9:beta&gamma;10:3normal
        icode = 0   * 0:standard genetic code; 1:
           mammalian mt; 2-10:see below
        Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
```

```
            4:all

    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
       estimated
        kappa = .3   * initial or fixed kappa
    fix_omega = 0   * 1: omega or omega_1 fixed, 0:
       estimate
        omega = 1.5  * initial or fixed omega, for codons
            or codon-based AAs
        ncatG = 10   * # of categories in the dG or AdG
           models of rates

        getSE = 0   * 0: don't want them, 1: want S.E.s
           of estimates
 RateAncestor = 0   * (0,1,2): rates (alpha>0) or
    ancestral states (1 or 2)

   Small_Diff = .45e-6
    cleandata = 1  * remove sites with ambiguity data (1:
       yes, 0:no)?
  fix_blength = 0  * 0: ignore, -1: random, 1: initial,
     2: fixed
```

B.3 Branch-site tests

B.3.1 Null model

```
        seqfile = branchsite_test/desat1_all_codons_4paml
           .phy    * sequence data file name
     treefile = branchsite_test/two_ratios_leadingbranch.
        tree   * tree structure file name
```

```
      outfile = branchsite_test/
         two_ratios_leadingbranch_alt.paml  * main result
          file name
        noisy = 3   * 0,1,2,3,9: how much rubbish on the
           screen
      verbose = 0   * 1: detailed output, 0: concise
         output
      runmode = 0   * 0: user tree;  1: semi-automatic;
         2: automatic
                    * 3: StepwiseAddition; (4,5):
                       PerturbationNNI; -2: pairwise

        seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
       codon table
        clock = 0   * 0: no clock, unrooted tree, 1:
           clock, rooted tree
       aaDist = 0   * 0:equal, +:geometric; -:linear,
          {1-5:G1974,Miyata,c,p,v}
        model = 2

      NSsites = 2
                    * 0:one w; 1:NearlyNeutral; 2:
                       PositiveSelection; 3:discrete;
                    * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                       beta&w;9:beta&gamma;10:3normal
        icode = 0   * 0:standard genetic code; 1:
           mammalian mt; 2-10:see below
        Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
            4:all
```

```
    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
       estimated
        kappa = .3   * initial or fixed kappa
    fix_omega = 1   * 1: omega or omega_1 fixed, 0:
       estimate
        omega = 1  * initial or fixed omega, for codons
           or codon-based AAs
        ncatG = 10   * # of categories in the dG or AdG
           models of rates

        getSE = 0   * 0: don't want them, 1: want S.E.s
           of estimates
 RateAncestor = 0   * (0,1,2): rates (alpha>0) or
    ancestral states (1 or 2)

   Small_Diff = .45e-6
    cleandata = 1  * remove sites with ambiguity data (1:
       yes, 0:no)?
  fix_blength = 0  * 0: ignore, -1: random, 1: initial,
     2: fixed
```

B.3.2 Alternative model

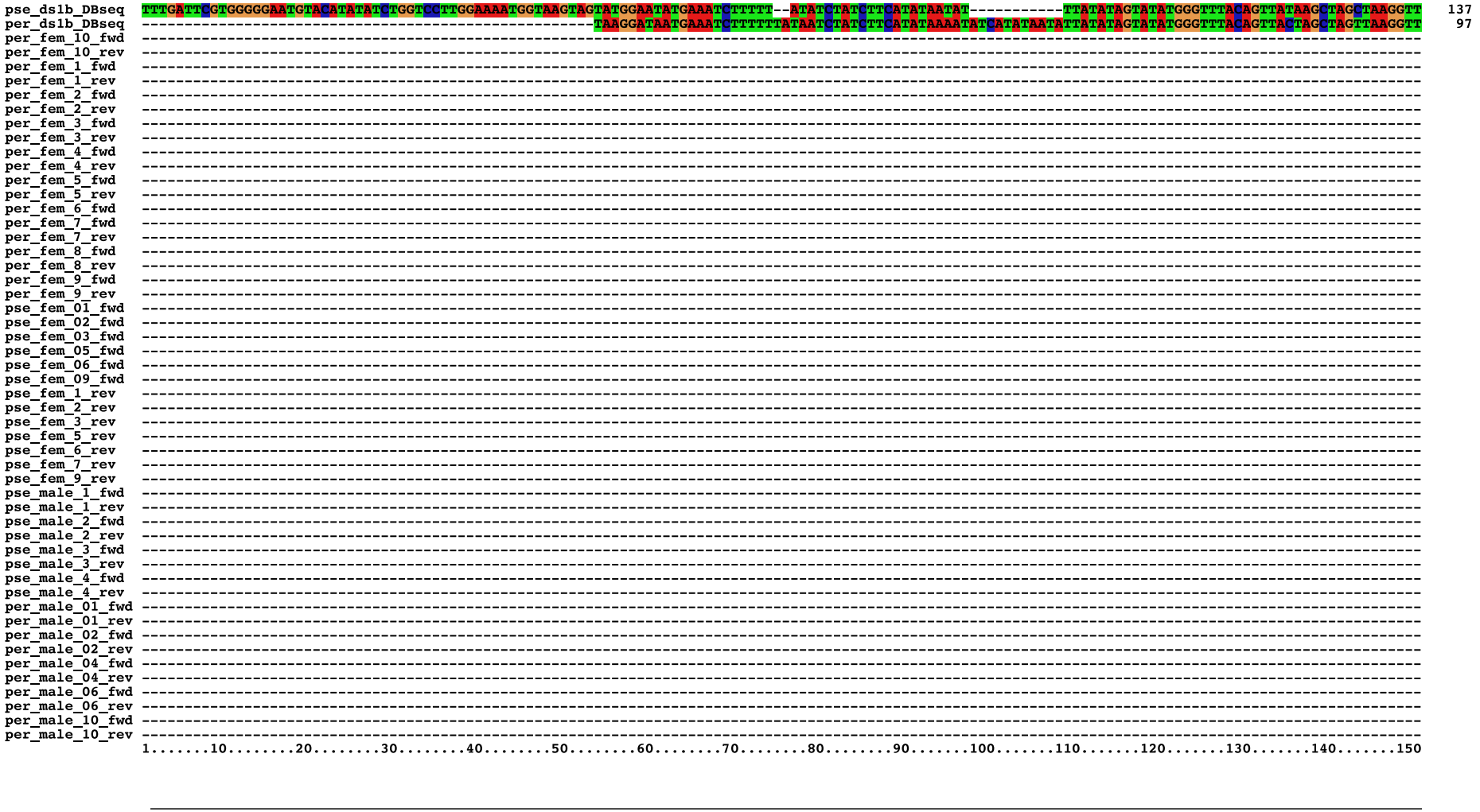
```
        seqfile = branchsite_test/desat1_all_codons_4paml
           .phy    * sequence data file name
     treefile = branchsite_test/two_ratios_leadingbranch.
        tree   * tree structure file name

      outfile = branchsite_test/
         two_ratios_leadingbranch_alt.paml   * main
         result file name
        noisy = 3   * 0,1,2,3,9: how much rubbish on the
```

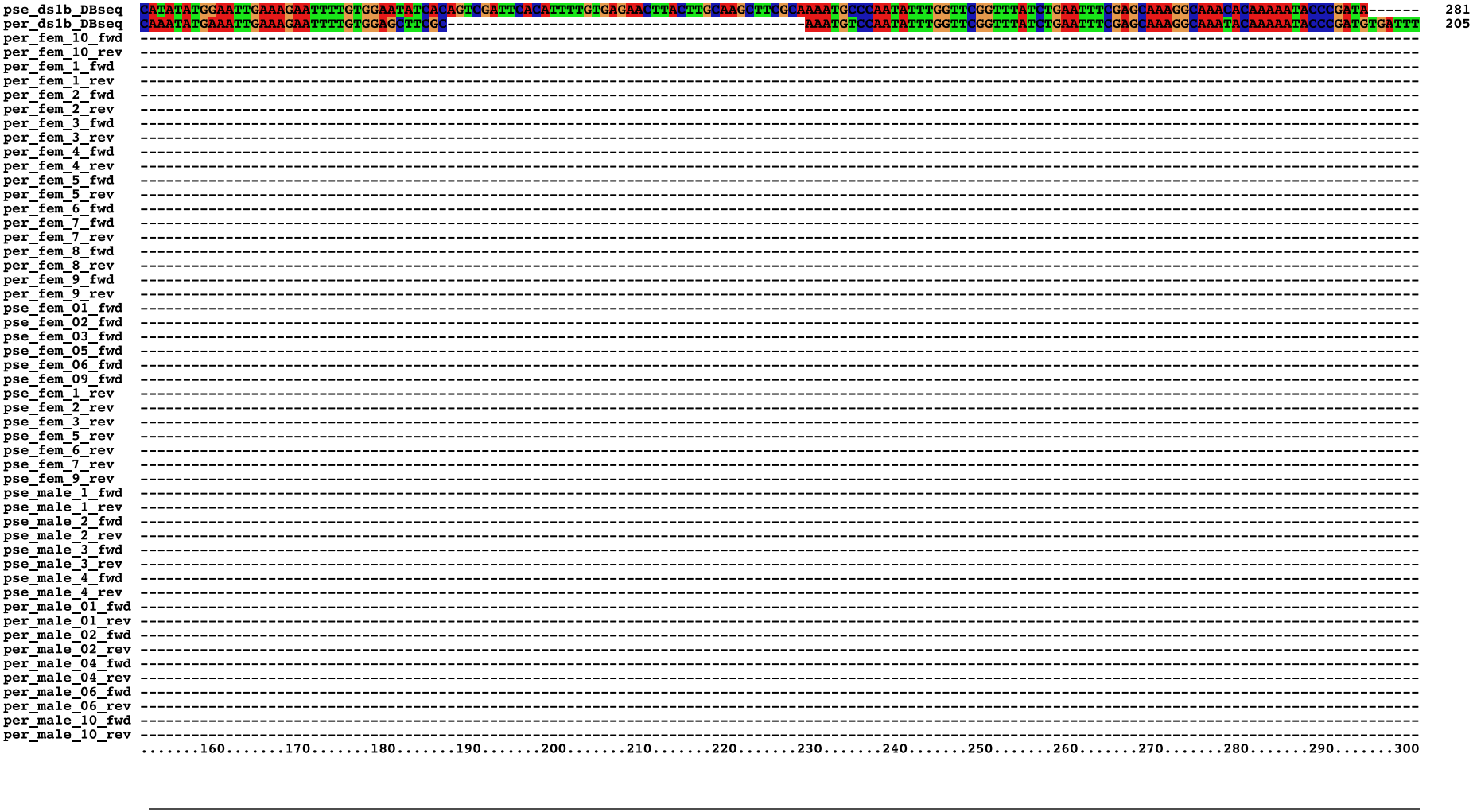
```
        screen
    verbose = 0   * 1: detailed output, 0: concise
       output
    runmode = 0   * 0: user tree;  1: semi-automatic;
       2: automatic
                  * 3: StepwiseAddition; (4,5):
                     PerturbationNNI; -2: pairwise

    seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
  CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:
     codon table
      clock = 0   * 0: no clock, unrooted tree, 1:
         clock, rooted tree
     aaDist = 0   * 0:equal, +:geometric; -:linear,
        {1-5:G1974,Miyata,c,p,v}
      model = 2

    NSsites = 2
                  * 0:one w; 1:NearlyNeutral; 2:
                     PositiveSelection; 3:discrete;
                  * 4:freqs; 5:gamma;6:2gamma;7:beta;8:
                     beta&w;9:beta&gamma;10:3normal
      icode = 0   * 0:standard genetic code; 1:
         mammalian mt; 2-10:see below
      Mgene = 0   * 0:rates, 1:separate; 2:pi, 3:kappa,
          4:all

  fix_kappa = 0   * 1: kappa fixed, 0: kappa to be
     estimated
      kappa = .3   * initial or fixed kappa
  fix_omega = 0   * 1: omega or omega_1 fixed, 0:
```

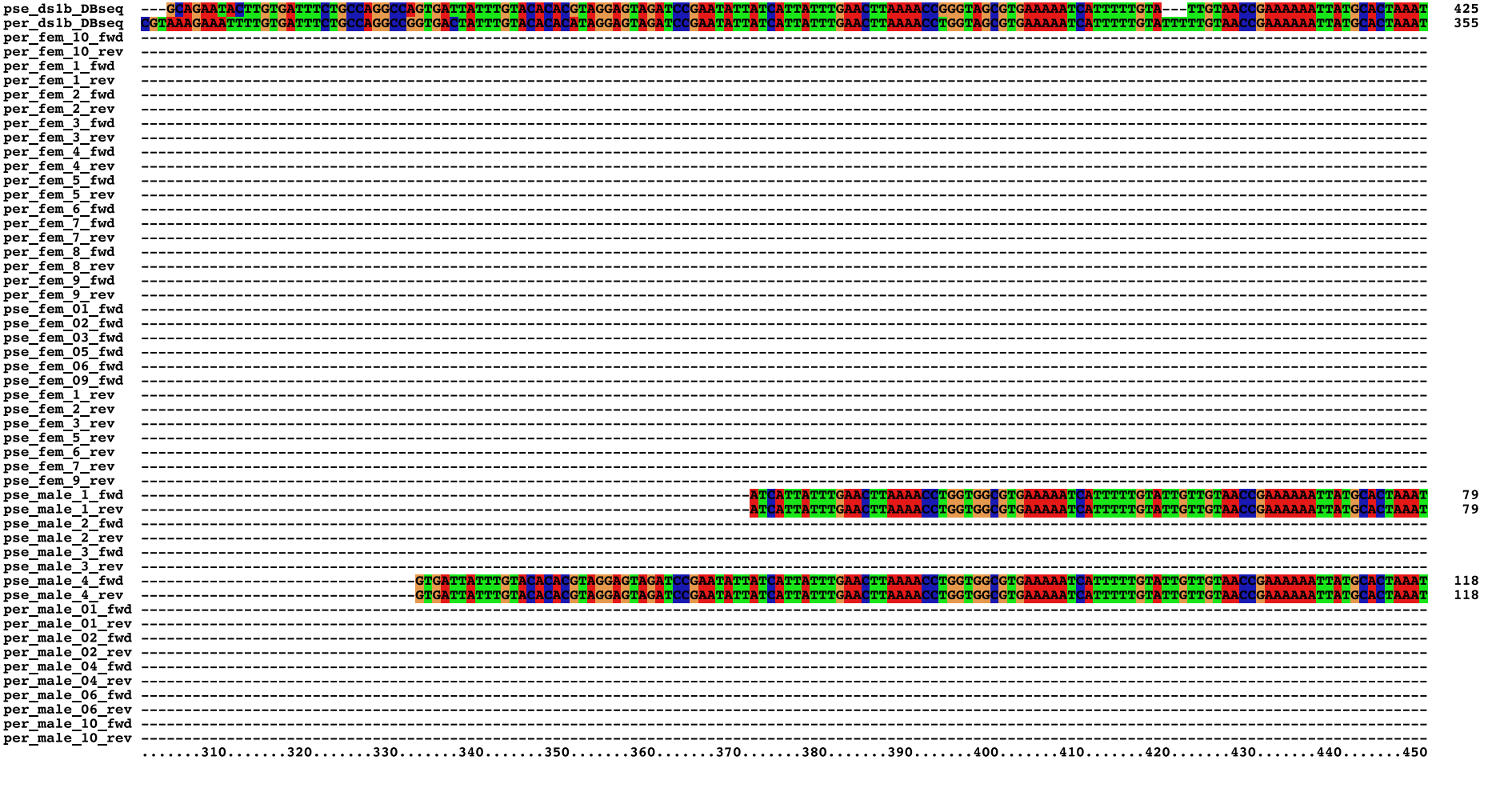
```
      estimate
       omega = 1.5  * initial or fixed omega, for codons
           or codon-based AAs
       ncatG = 10   * # of categories in the dG or AdG
          models of rates

       getSE = 0    * 0: don't want them, 1: want S.E.s
          of estimates
RateAncestor = 0    * (0,1,2): rates (alpha>0) or
   ancestral states (1 or 2)

  Small_Diff = .45e-6
   cleandata = 1  * remove sites with ambiguity data (1:
      yes, 0:no)?
 fix_blength = 0  * 0: ignore, -1: random, 1: initial,
    2: fixed
```

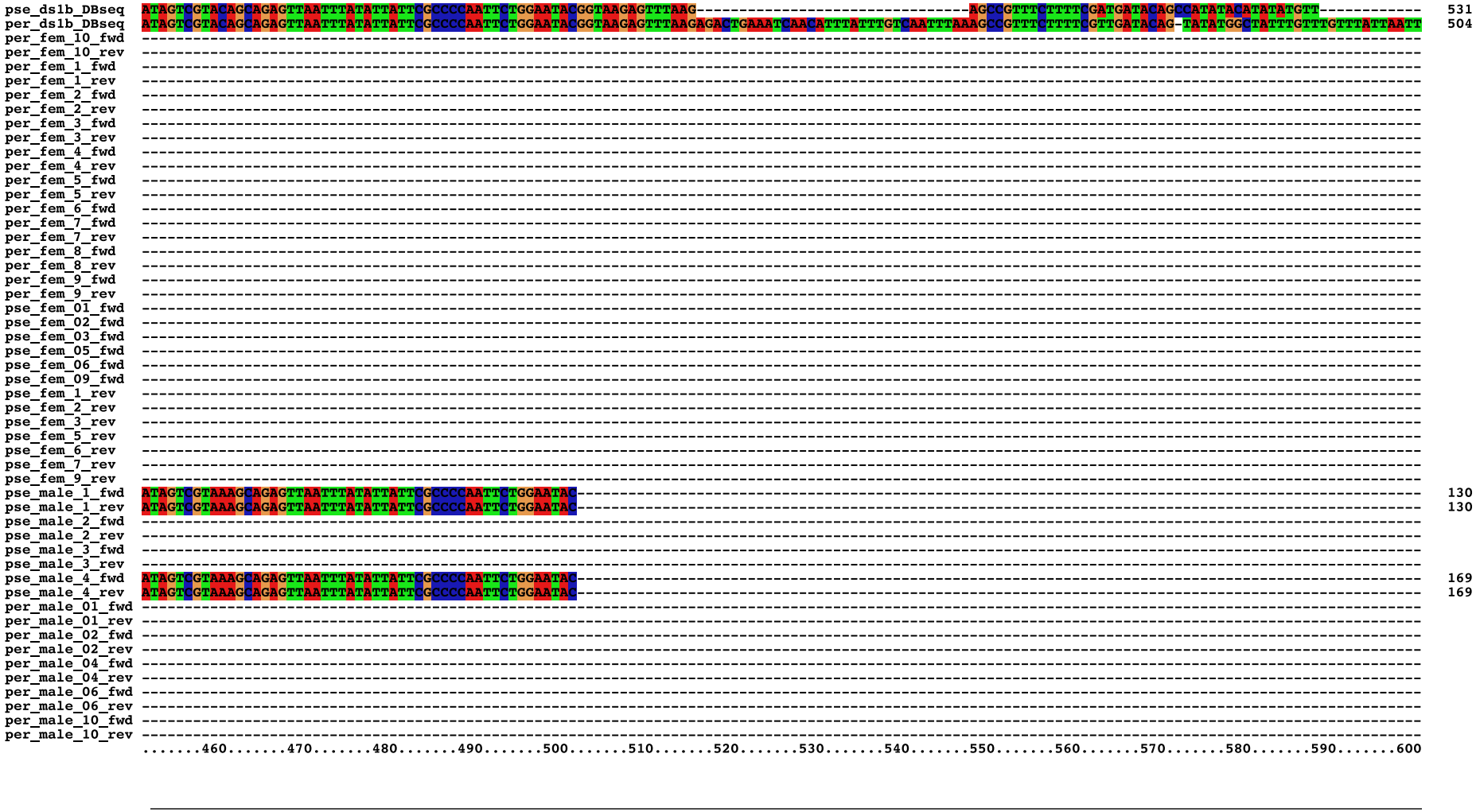
Appendix C

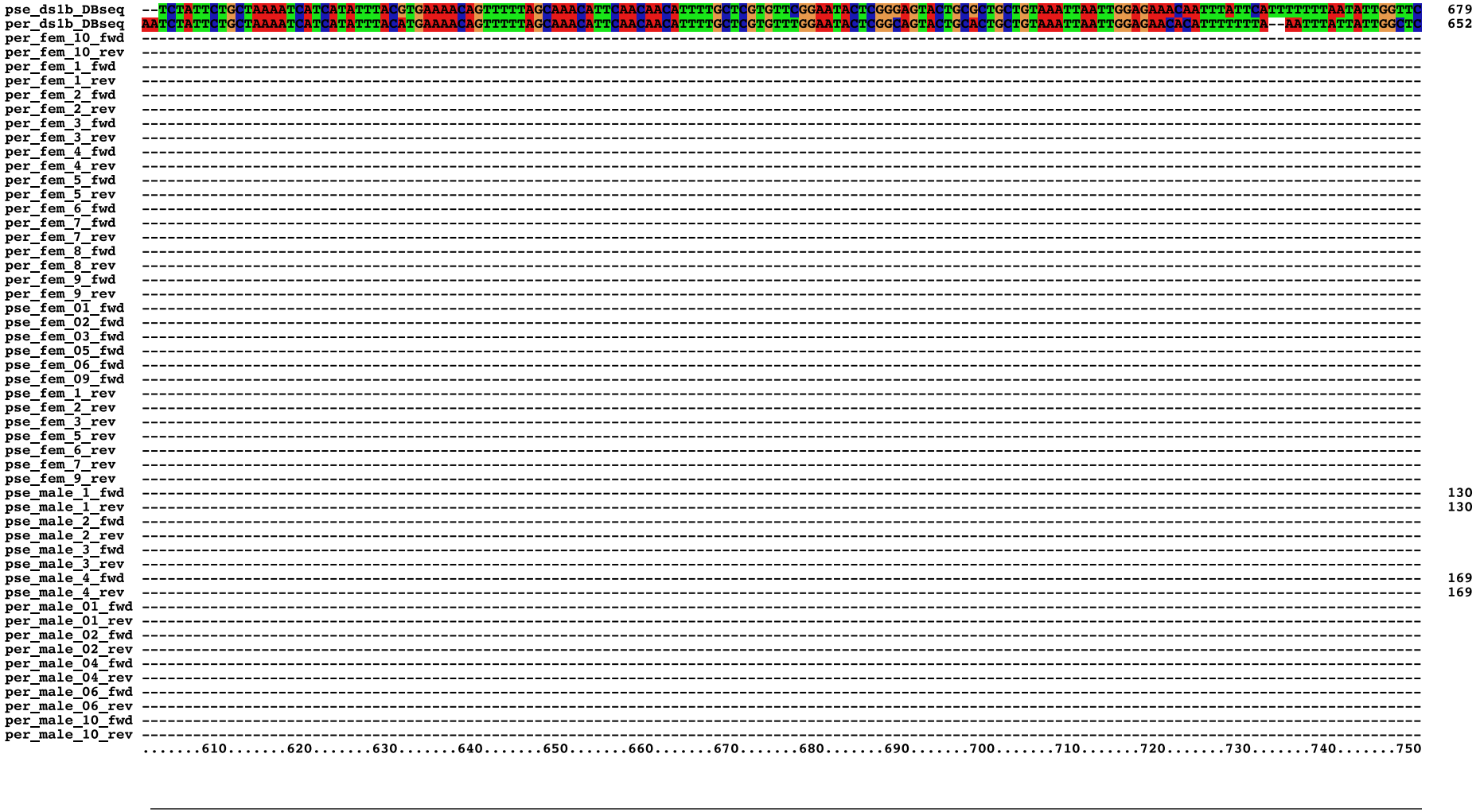
Alignment of *desat1b* RACE products

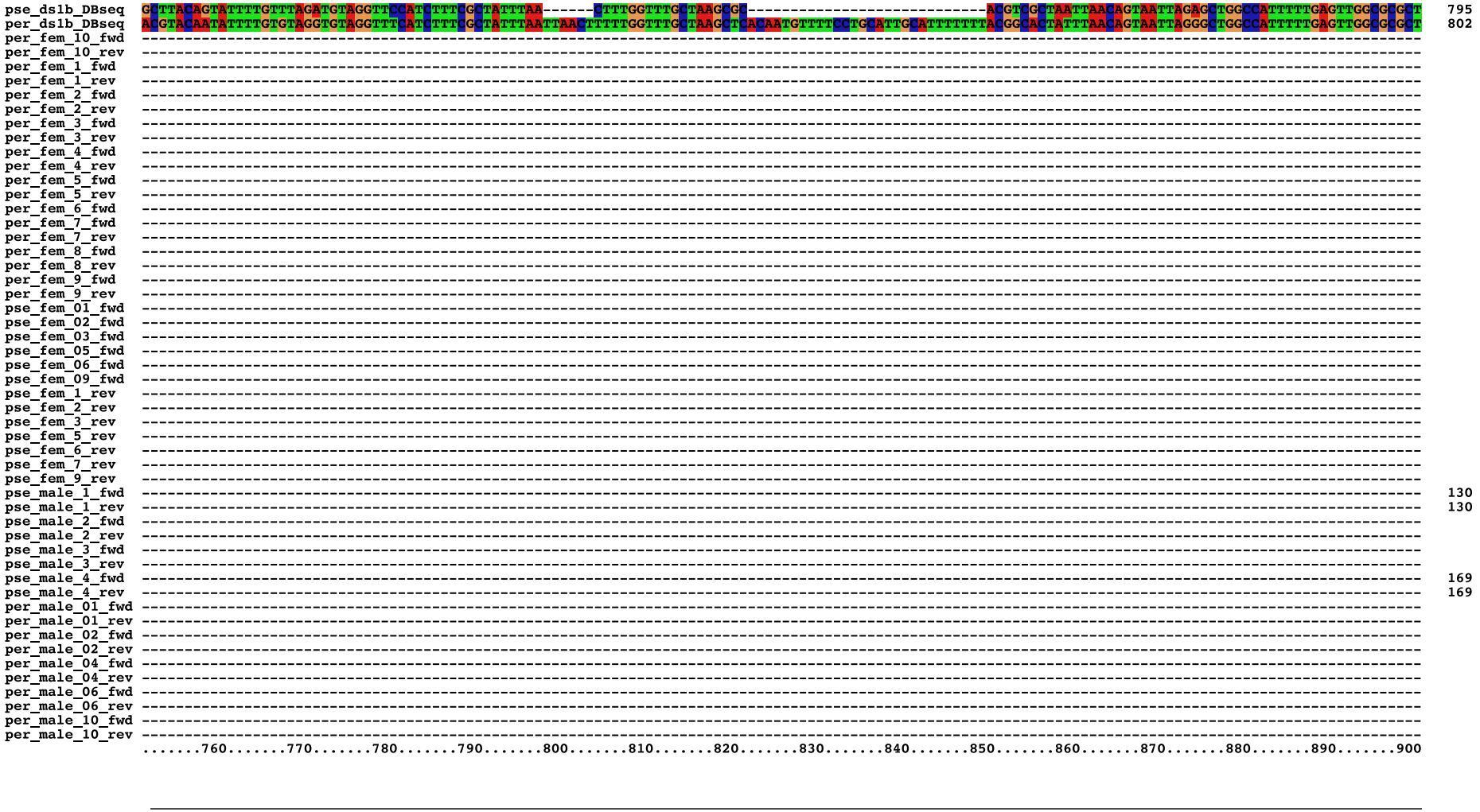

pse_ds1b_DBseq TTTGATTCGTGGGGGAATGTACATATATCTGGTCCTTGGAAAATGGTAAGTAGTATGGAATATGAAATCTTTTT--ATATCTATCTTCATATAATAT----------TTATATAGTATATGGGTTTACAGTTATAAGTTAGTTAAGGTT 137
per_ds1b_DBseq ----------TAAGGATAATGAAATCTTTTTTATAATCTATCTTCATATAAAATATCATATAATATTATATAGTATATGGGTTTACAGTTACTAGTTAGTTAAGGTT 97
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150

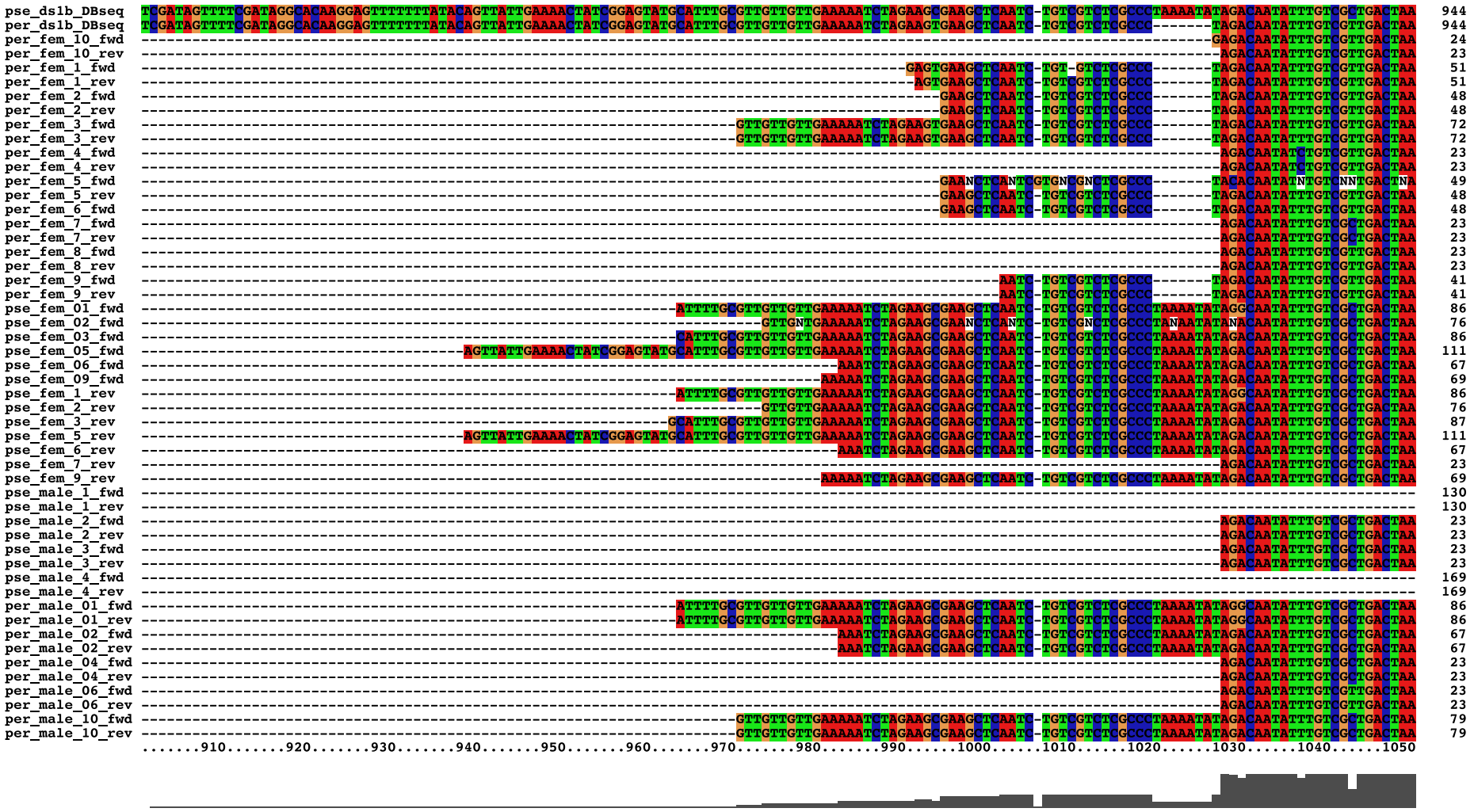

pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
281
205
.....160.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....290.....300

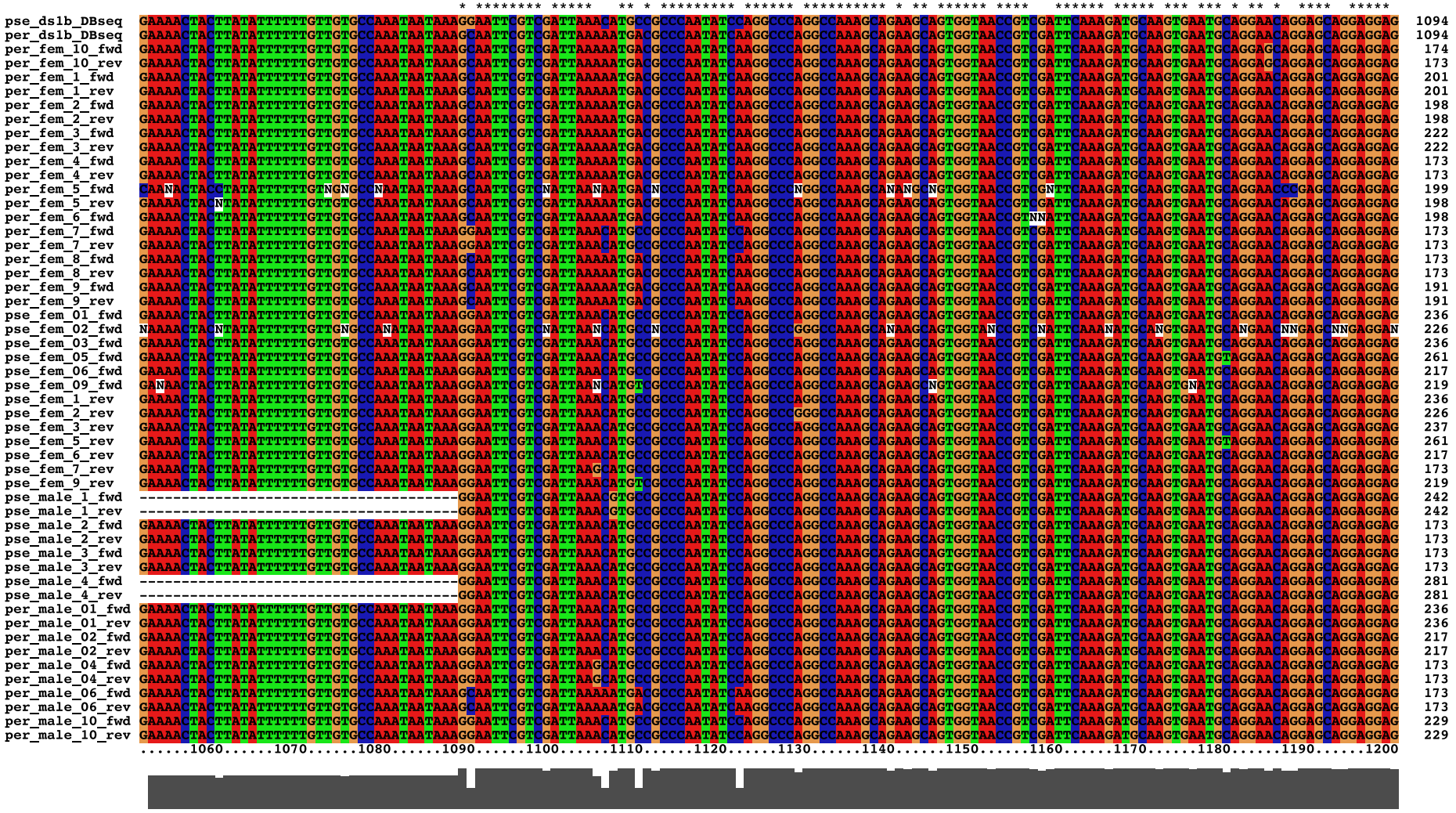

pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
425
355
79
79
118
118
.....310.....320.....330.....340.....350.....360.....370.....380.....390.....400.....410.....420.....430.....440.....450

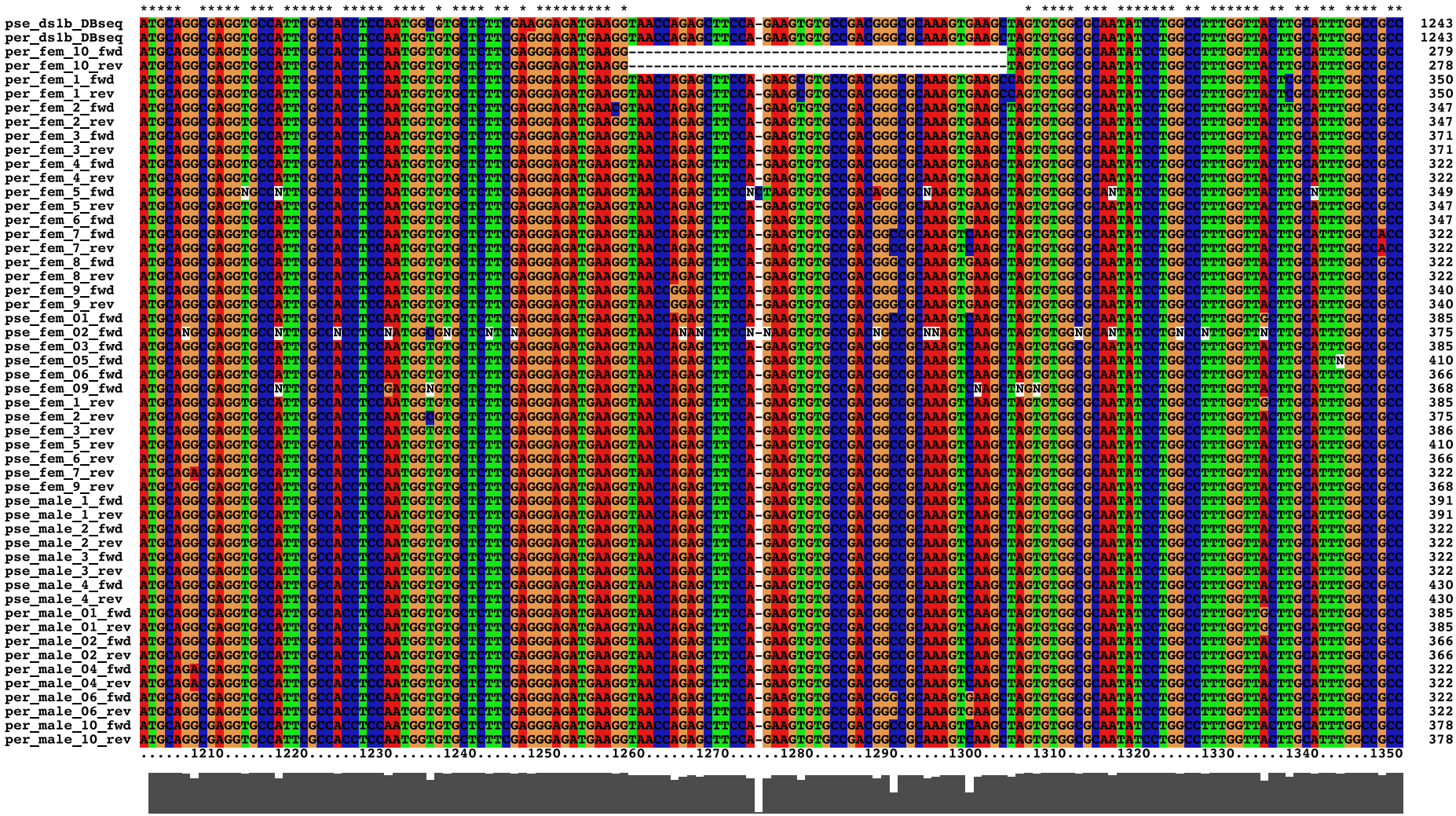

pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
531
504
130
130
169
169
.....460.....470.....480.....490.....500.....510.....520.....530.....540.....550.....560.....570.....580.....590.....600

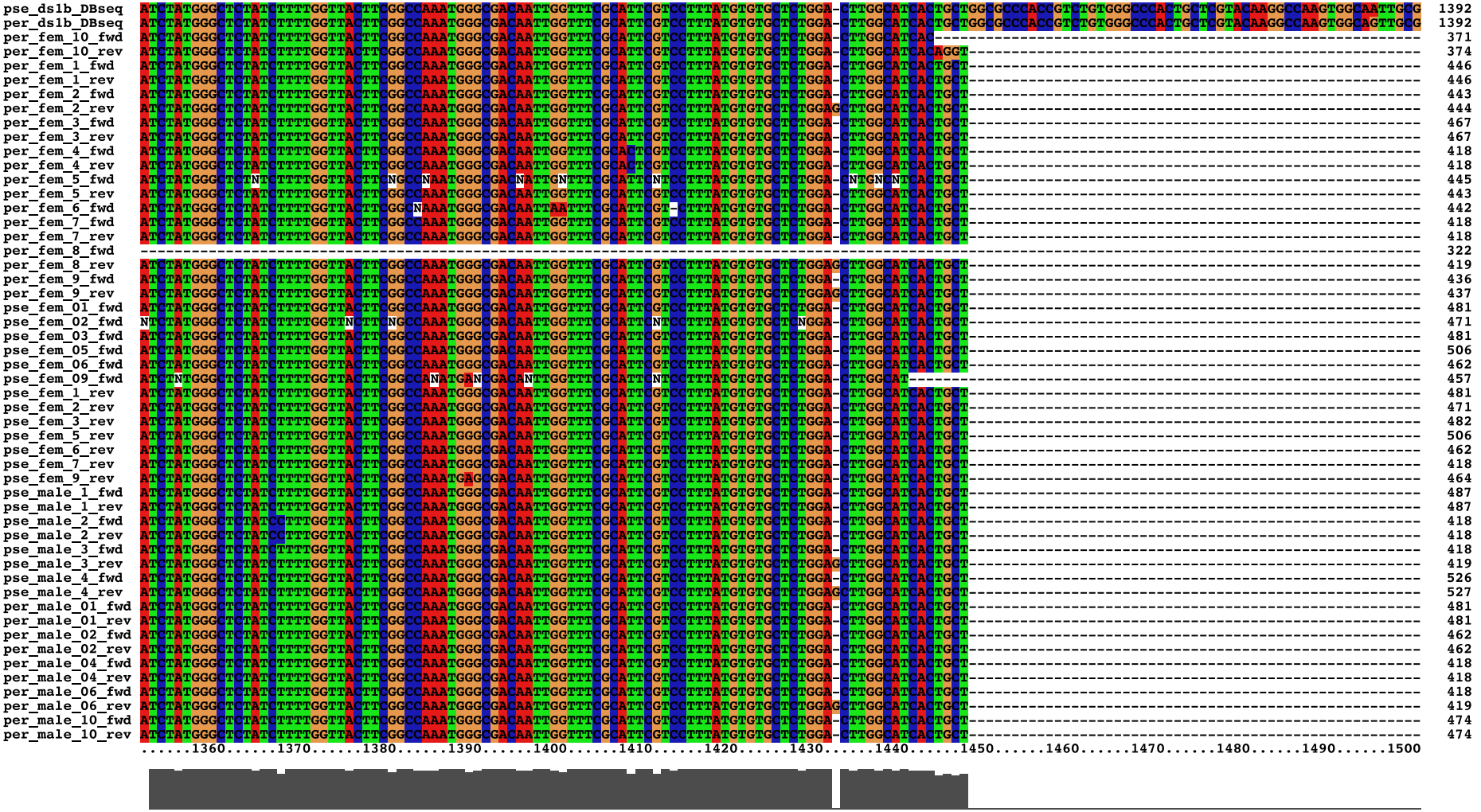

pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
679
652
130
130
169
169
.....610.....620.....630.....640.....650.....660.....670.....680.....690.....700.....710.....720.....730.....740.....750


pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
795
802
130
130
169
169
.....760.....770.....780.....790.....800.....810.....820.....830.....840.....850.....860.....870.....880.....890.....900


pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
.....910.....920.....930.....940.....950.....960.....970.....980.....990.....1000.....1010.....1020.....1030.....1040.....1050


pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
1094
1094
174
173
201
201
198
198
222
222
173
173
199
198
198
173
173
173
191
191
236
226
236
261
217
219
236
226
237
261
217
173
219
242
242
173
173
173
173
281
281
236
236
217
173
173
229
229
.....1060.....1070.....1080.....1090.....1100.....1110.....1120.....1130.....1140.....1150.....1160.....1170.....1180.....1190.....1200


pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350


pse_ds1b_DBseq
per_ds1b_DBseq
per_fem_10_fwd
per_fem_10_rev
per_fem_1_fwd
per_fem_1_rev
per_fem_2_fwd
per_fem_2_rev
per_fem_3_fwd
per_fem_3_rev
per_fem_4_fwd
per_fem_4_rev
per_fem_5_fwd
per_fem_5_rev
per_fem_6_fwd
per_fem_7_fwd
per_fem_7_rev
per_fem_8_fwd
per_fem_8_rev
per_fem_9_fwd
per_fem_9_rev
pse_fem_01_fwd
pse_fem_02_fwd
pse_fem_03_fwd
pse_fem_05_fwd
pse_fem_06_fwd
pse_fem_09_fwd
pse_fem_1_rev
pse_fem_2_rev
pse_fem_3_rev
pse_fem_5_rev
pse_fem_6_rev
pse_fem_7_rev
pse_fem_9_rev
pse_male_1_fwd
pse_male_1_rev
pse_male_2_fwd
pse_male_2_rev
pse_male_3_fwd
pse_male_3_rev
pse_male_4_fwd
pse_male_4_rev
per_male_01_fwd
per_male_01_rev
per_male_02_fwd
per_male_02_rev
per_male_04_fwd
per_male_04_rev
per_male_06_fwd
per_male_06_rev
per_male_10_fwd
per_male_10_rev
1392
1392
371
374
446
446
443
444
467
467
418
418
445
443
442
418
418
322
419
436
437
481
471
481
506
462
457
481
471
482
506
462
418
464
487
487
418
418
418
526
527
481
481
462
462
418
418
419
474
474
.....1360.....1370.....1380.....1390.....1400.....1410.....1420.....1430.....1440.....1450.....1460.....1470.....1480.....1490.....1500

Bibliography

Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R. and Engels, W. R. (1991). Targeted gene replacement in *Drosophila* via P element-induced gap repair. *Science*, **253**:1110–1117.